

第8章 Internet及其应用

Internet概述



Internet基本服务与应用



网络信息检索技术



网络信息安全技术

8.1 Internet概述

8.1.1 网络的网络

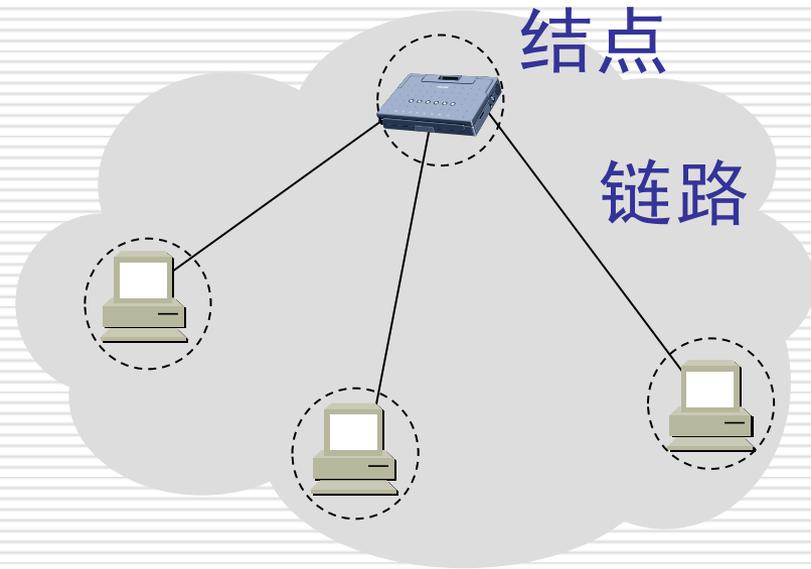
- 起源于美国的**Internet**现已发展成为世界上最大的国际性计算机互联网
 - **网络(network)**由若干**结点(node)**和连接这些结点的**链路(link)**组成。
 - 互联网是“**网络的网络**”(network of networks)。
 - 连接在**Internet**上的计算机都称为**主机(host)**。
-

网络与Internet

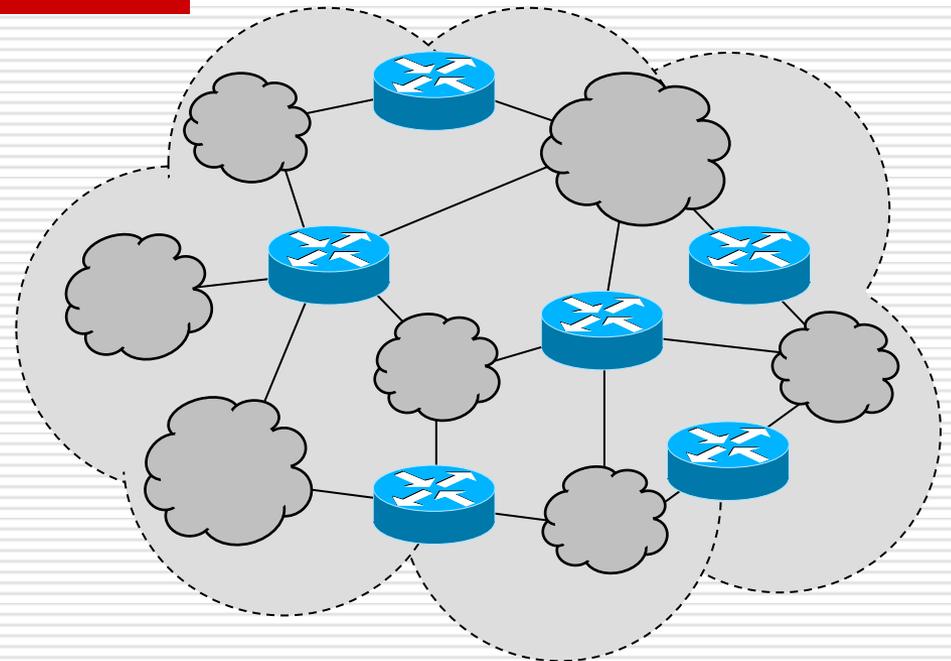
- 网络把许多计算机连接在一起。
 - **Internet**则把许多网络连接在一起。
-

网络

互联网（网络的网络）

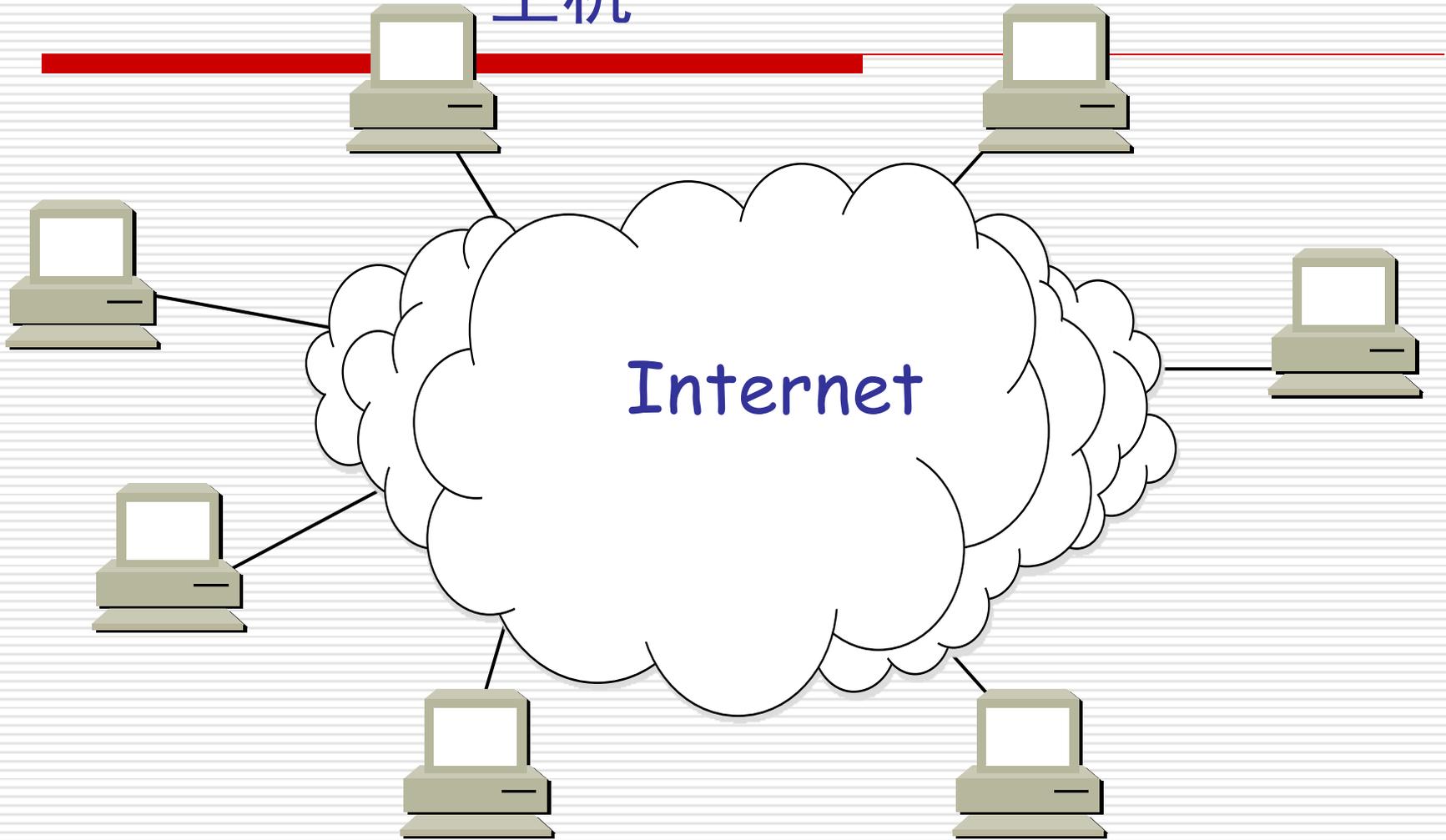


(a)



(b)

主机



8.1.2 Internet发展的三个阶段

1

从单个网络 **ARPANET** 向互联网的发展

1983 年 **TCP/IP** 协议成为 **ARPANET** 上的标准协议。一般认为1983 年为**Internet**的诞生时间。

2

建成了三级结构的**Internet**

三级计算机网络，分为主干网、地区网和校园网(或企业网)

3

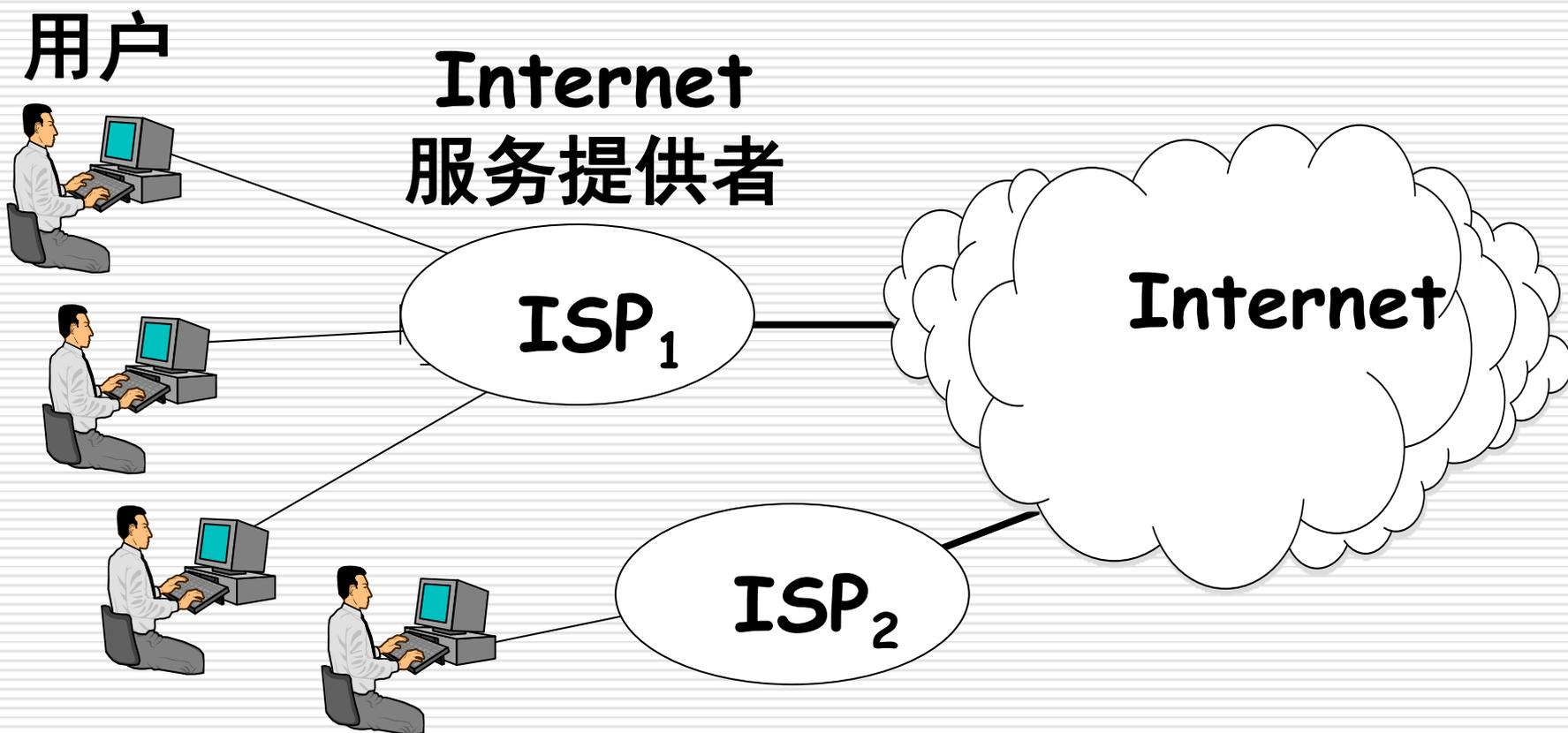
逐渐形成多层次 **ISP** 结构的**Internet**

出现了**Internet**服务提供商 **ISP** (Internet Service Provider)

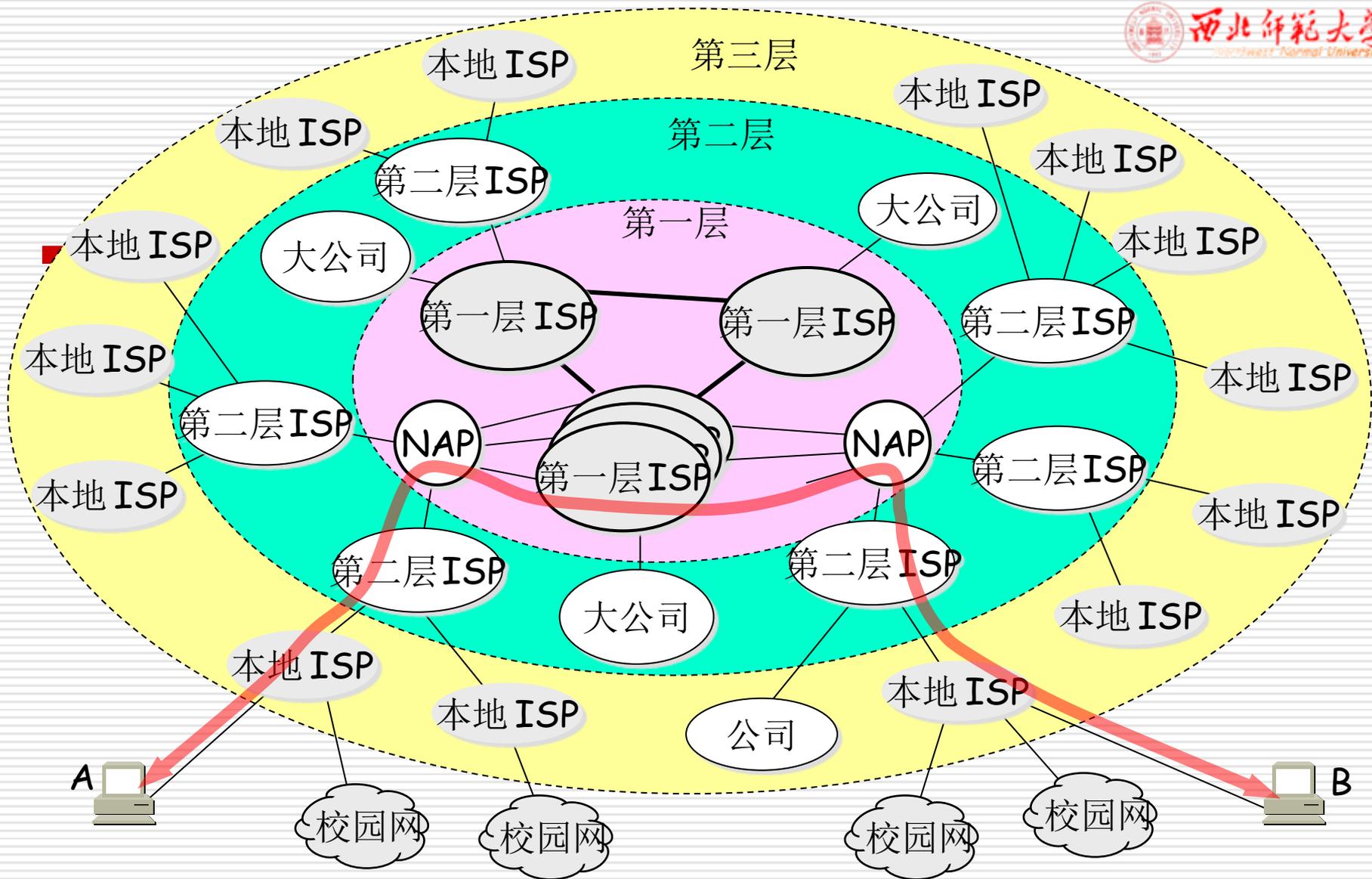
internet 和 Internet 的区别

- **internet**（互联网或互连网）是一个通用名词，泛指由多个计算机网络互连而成的网络。
 - **Internet**（因特网）是一个专用名词，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，采用 **TCP/IP** 协议族作为通信的规则，且其前身是美国的 **ARPANET**。
-

用户通过 ISP 上网



根据提供服务的覆盖面积大小以及所拥有的IP 地址数目的不同，ISP 也分成为不同的层次。

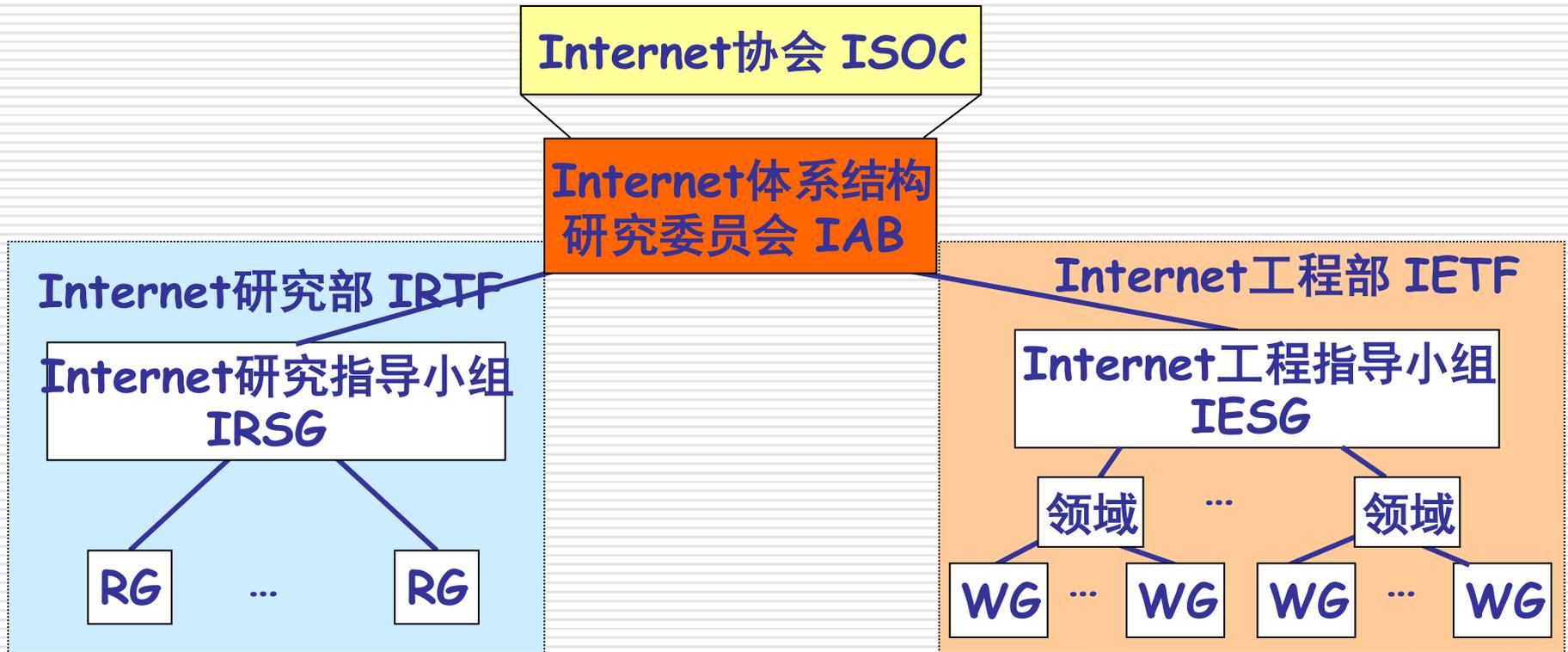


主机A → 本地 ISP → 第二层 ISP → NAP → 第一层 ISP → NAP → 第二层 ISP → 本地 ISP → 主机B

万维网 WWW 的问世

- **Internet**已经成为世界上规模最大和增长速率最快的计算机网络，没有人能够准确说出**Internet**究竟有多大。
 - **Internet**的迅猛发展始于 20 世纪 90 年代。由欧洲原子核研究组织 **CERN** 开发的万维网 **WWW (World Wide Web)**被广泛使用在**Internet**上，方便了广大非网络专业人员对网络的使用，成为**Internet**的这种指数级增长的主要驱动力。
-

8.1.3 Internet的标准化工作



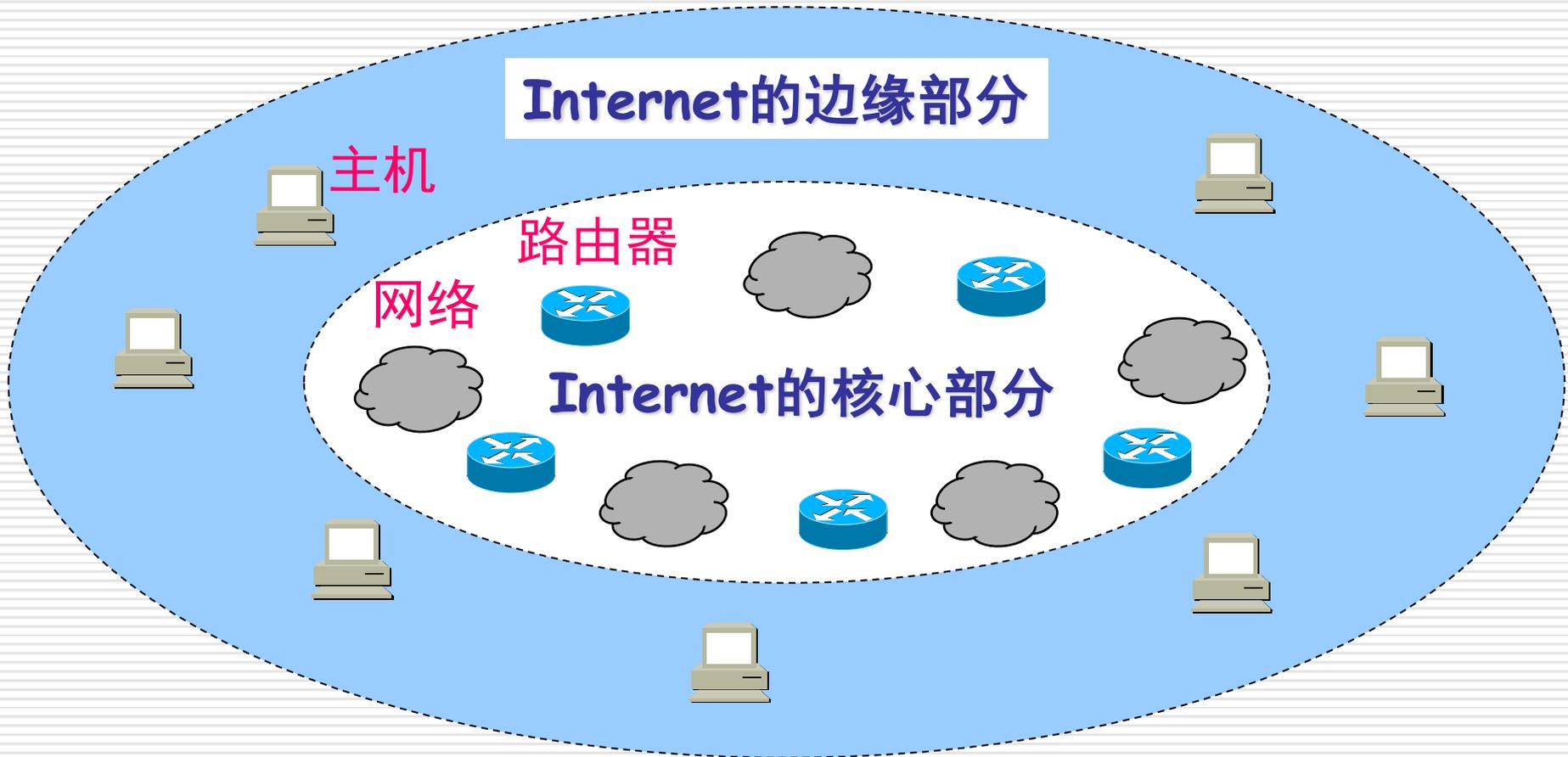
8.1.4 Internet的组成

从Internet的工作方式看，可分两大块：

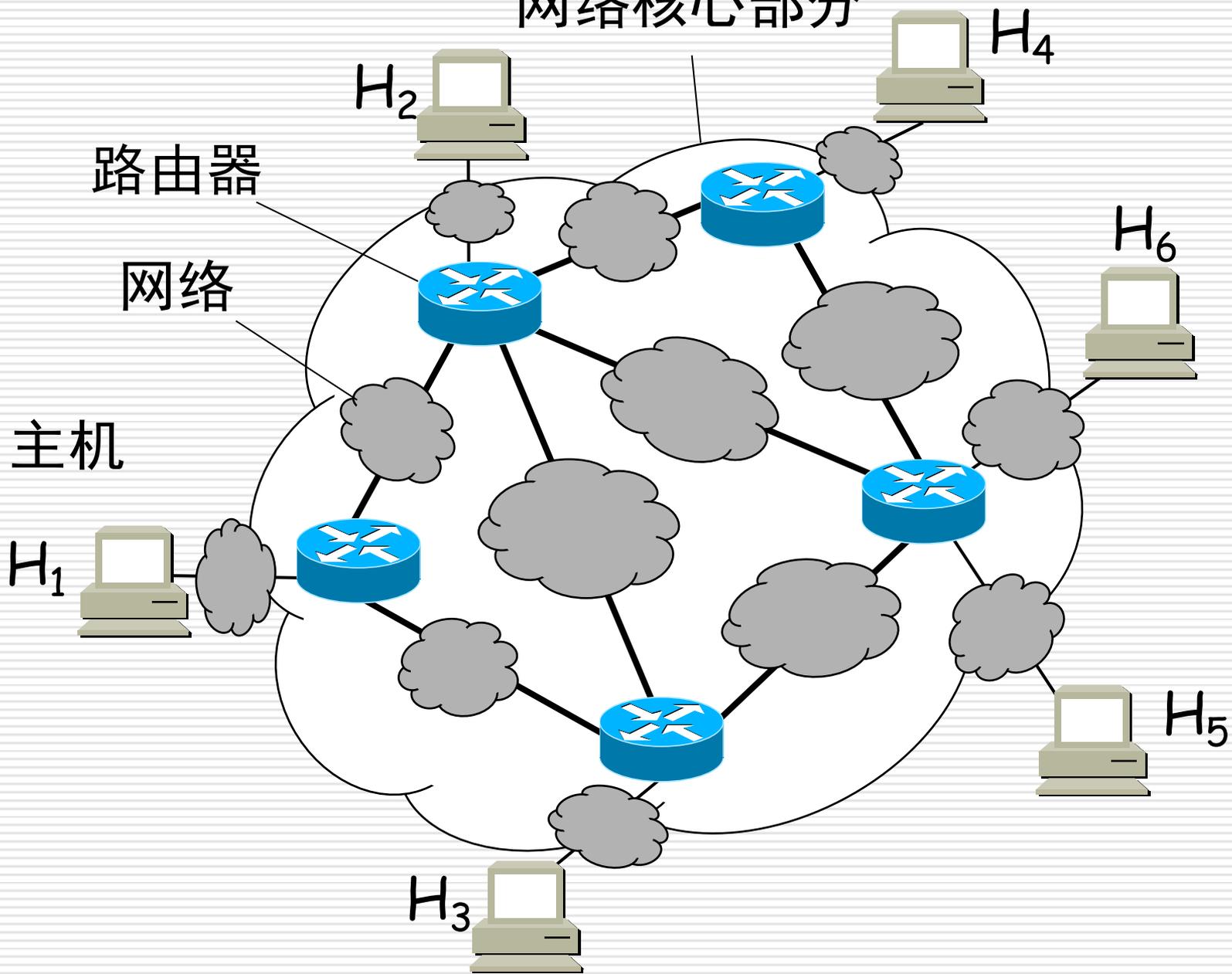
(1) **边缘部分** 由所有连接在Internet上的主机组成。这部分是用户直接使用的，用来进行通信（传送数据、音频或视频）和资源共享。

(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）。

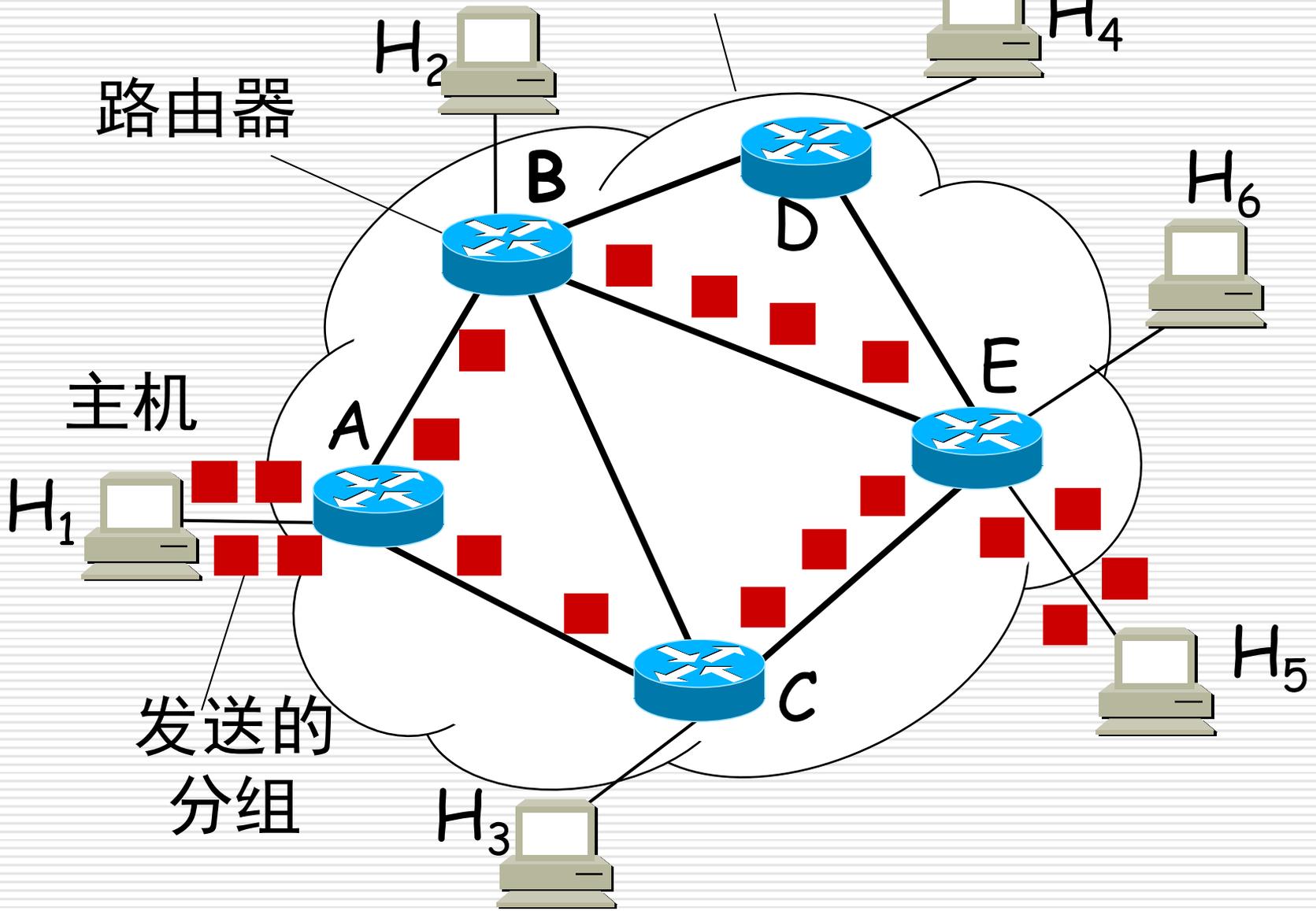
Internet的边缘部分与核心部分



网络核心部分



网络核心部分



路由器

主机

发送的
分组

H₂

H₄

H₆

H₁

H₅

H₃

A

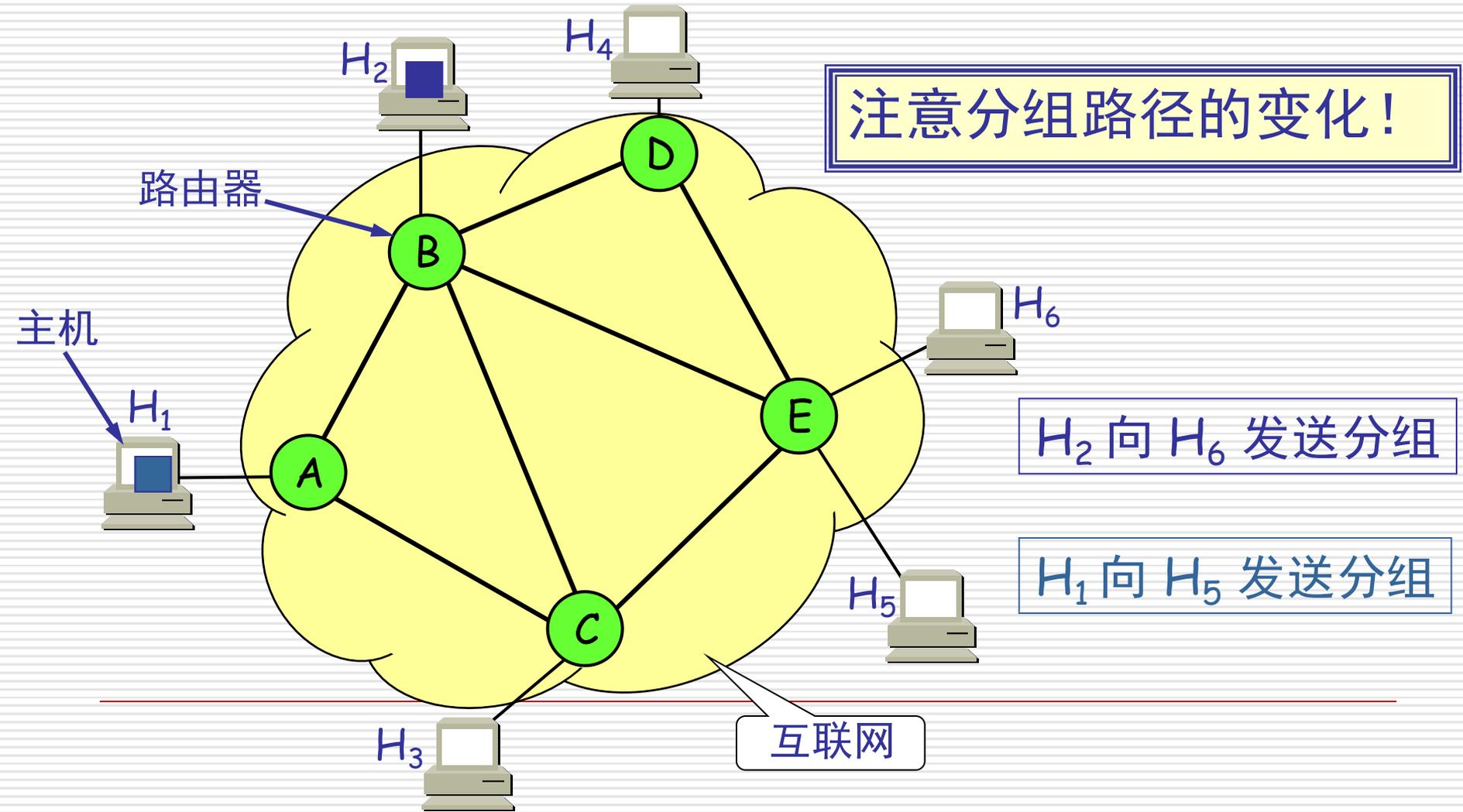
B

C

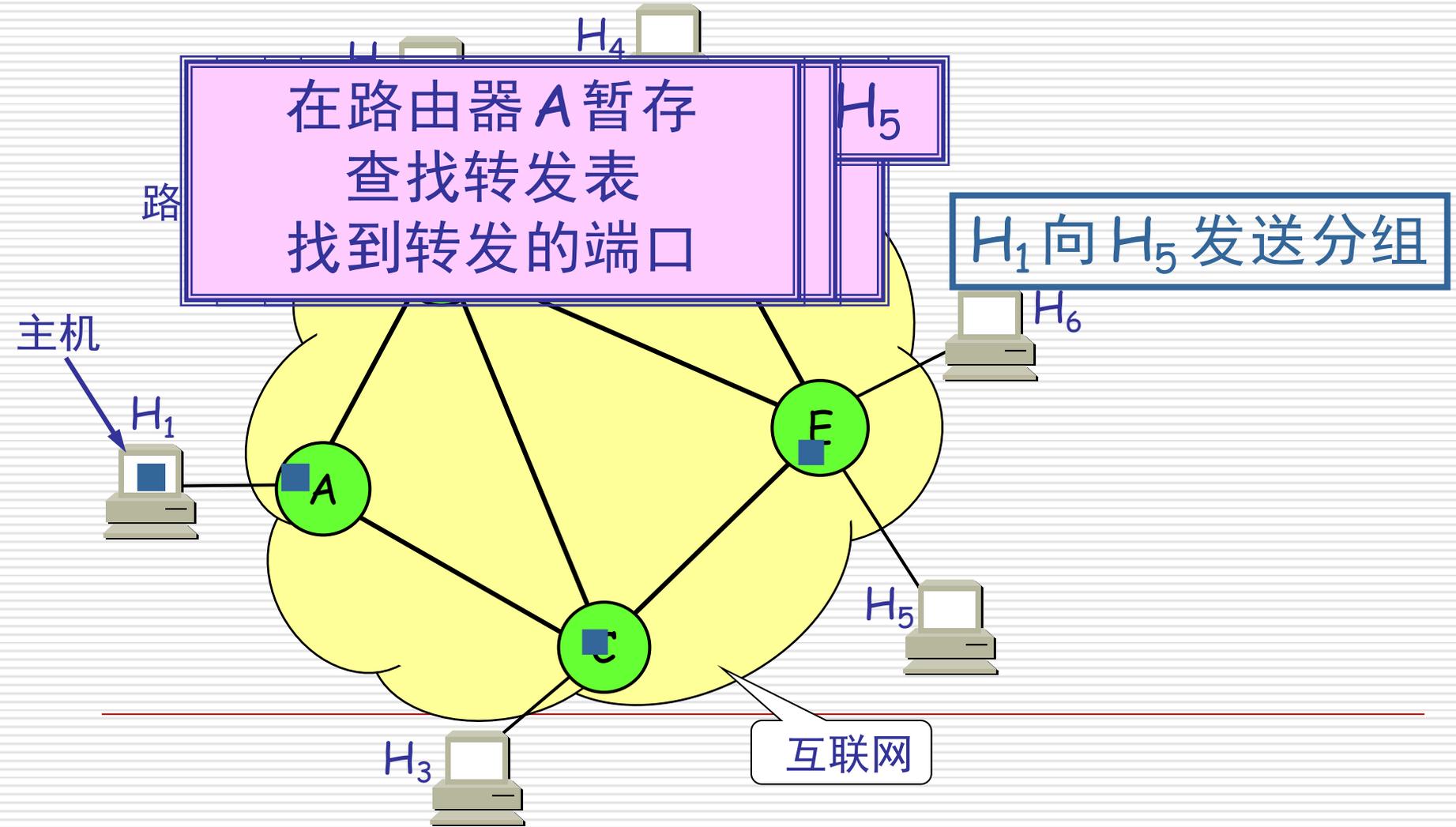
D

E

分组交换网的示意图



注意分组的存储转发过程



主机和路由器的作用不同

- 主机是为用户进行信息处理的，并向网络发送分组，从网络接收分组。
 - 路由器对分组进行存储转发，最后把分组交付目的主机。
-

ARPANET的成功使 计算机网络概念发生了根本变化

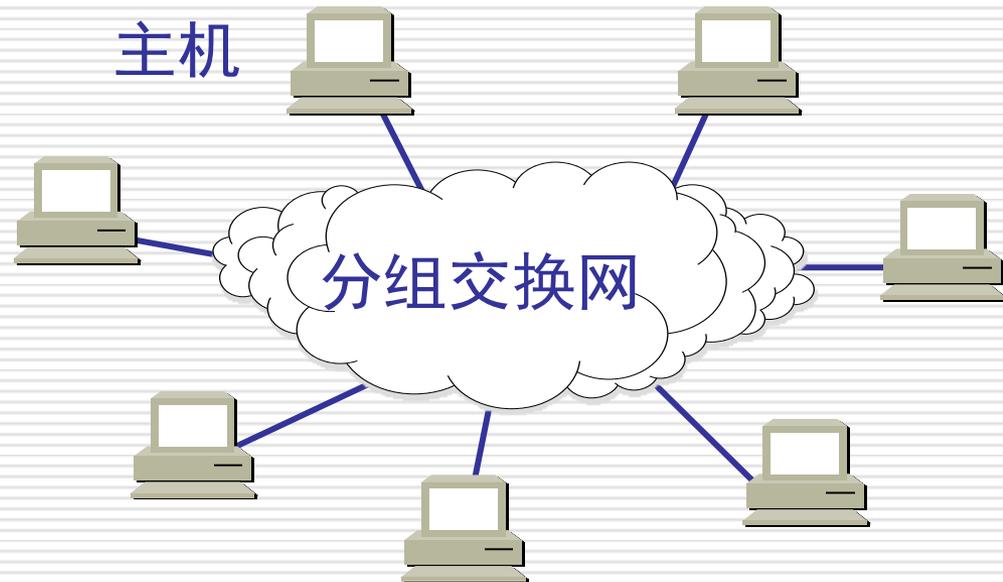
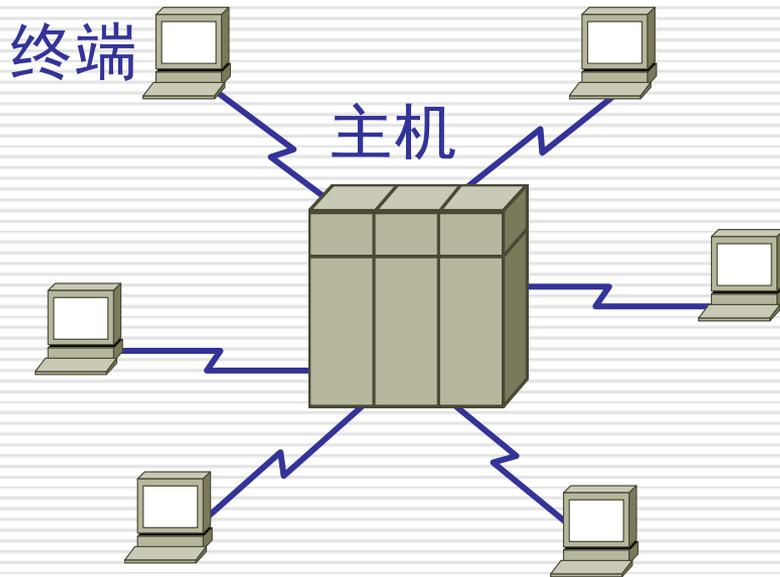
- 早期的面向终端的计算机网络是以**单个主机为中心的星形网**
 - 各终端通过通信线路共享昂贵的中心主机的硬件和软件资源。
 - 分组交换网则是以**网络为中心**，主机都处在网络的外围。
 - 用户通过分组交换网可共享连接在网络上的许多硬件和各种丰富的软件资源。
-

从主机为中心到以网络为中心

以主机为中心



以分组交换网为中心



8.1.5 计算机网络在我国的发展

- (1) 中国公用计算机互联网 **CHINANET**
 - (2) 中国教育和科研计算机网 **CERNET**
 - (3) 中国科学技术网 **CSTNET**
 - (4) 中国联通互联网 **UNINET**
 - (5) 中国网通公用互联网 **CNCNET**
 - (6) 中国国际经济贸易互联网 **CIETNET**
 - (7) 中国移动互联网 **CMNET**
 - (8) 中国长城互联网 **CGWNET**（建设中）
 - (9) 中国卫星集团互联网 **CSNET**（建设中）
-

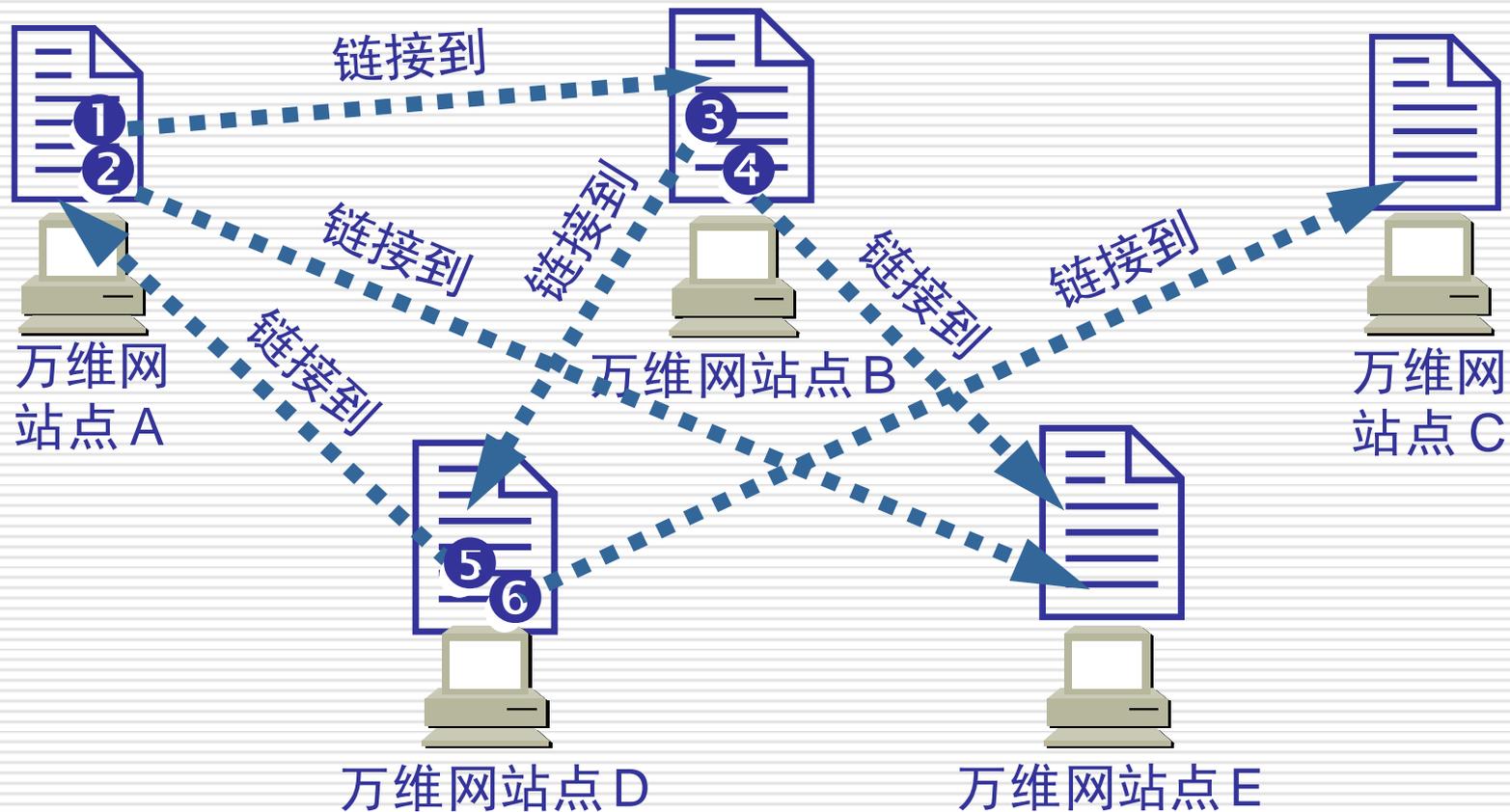
8.2 Internet基本服务与应用

8.2.1 万维网 WWW

1 万维网概述

- **万维网 WWW (World Wide Web)**并非某种特殊的计算机网络。
 - 万维网是一个大规模的、联机式的信息储藏所。
 - 万维网用链接的方法能非常方便地从**Internet**上的一个站点访问另一个站点，从而主动地按需获取丰富的信息。
 - 这种访问方式称为“**链接**”。
-

万维网提供分布式服务



超媒体与超文本

- 万维网是**分布式超媒体(hypermedia)**系统，是**超文本(hypertext)**系统的扩充。
 - 一个超文本由多个信息源链接成。利用一个链接可使用户找到另一个文档。这些文档可以位于世界上任何一个接在**Internet**上的超文本系统中。超文本是万维网的基础。
 - 超媒体与超文本的区别是**文档内容**不同。超文本文档仅包含文本信息，而超媒体文档还包含其他表示方式的信息，如图形、图像、声音、动画，甚至活动视频图像。
-

万维网的工作方式

- 万维网以客户服务器方式工作。
 - **浏览器**就是在用户计算机上的万维网**客户程序**。万维网文档所驻留的计算机则运行**服务器程序**，因此这个计算机也称为**万维网服务器**。
 - 客户程序向服务器程序发出请求，服务器程序向客户程序送回客户所要的万维网文档。
 - 在一个客户程序主窗口上显示出的万维网文档称为**页面**(page)。
-

万维网必须解决的问题

(1) 怎样标志分布在整个Internet上的万维网文档？

- 使用**统一资源定位符 URL (Uniform Resource Locator)**来标志万维网上的各种文档。
 - 使每一个文档在整个**Internet**的范围内具有唯一的标识符 **URL**。
-

万维网必须解决的问题

(2) 用何协议实现万维网上各种超链的连接？

- 在万维网客户程序与万维网服务器程序之间进行交互所使用的协议，是**超文本传送协议 HTTP (HyperText Transfer Protocol)**。
 - **HTTP** 是一个应用层协议，它使用 **TCP** 连接进行可靠的传送。
-

万维网必须解决的问题

(3)怎样使各种万维网文档都能在Internet上的各种计算机上显示出来，同时使用户清楚地知道在什么地方存在着超链？

□ **超文本标记语言 HTML (HyperText Markup Language)**使得万维网页面的设计者可以很方便地用一个超链从本页面的某处链接到Internet上的任何一个万维网页面，并且能够在自己的计算机屏幕上将这些页面显示出来。

万维网必须解决的问题

(4) 怎样使用户能够很方便地找到所需的信息？

- 为了在万维网上方便地查找信息，用户可使用各种的搜索工具（即搜索引擎）。
-

2 统一资源定位符 URL

- URL 是对可以从**Internet**上得到的资源的位置和访问方法的一种简洁的表示。
 - URL 给资源的位置提供一种抽象的识别方法，并用这种方法给资源定位。
 - 只要能够对资源定位，系统就可以对资源进行各种操作，如存取、更新、替换和查找其属性。
 - URL 相当于一个文件名在网络范围的扩展。是与**Internet**相连的机器上任何可访问对象的指针。
-

URL 的一般形式

- 由以冒号隔开的两大部分组成，并且在 URL 中的字符对大写或小写没有要求。
- URL 的一般形式是：

<协议>://<主机>:<端口>/<路径>

ftp —— 文件传送协议 FTP

http —— 超文本传送协议 HTTP

News —— USENET 新闻

URL 的一般形式（续）

- 由以冒号隔开的两大部分组成，并且在 URL 中的字符对大写或小写没有要求。
- URL 的一般形式是：

〈协议〉://〈主机〉:〈端口〉/〈路径〉

〈主机〉是存放资源的主机
在Internet中的域名

URL 的一般形式（续）

- 由以冒号隔开的两大部分组成，并且在 URL 中的字符对大写或小写没有要求。
- URL 的一般形式是：

〈协议〉://〈主机〉:〈端口〉/〈路径〉

有时可省略

使用 HTTP 的 URL

- 使用 HTTP 的 URL 的一般形式

`http://<主机>:<端口>/<路径>`

↑
这表示使用 HTTP 协议

使用 HTTP 的 URL

- 使用 HTTP 的 URL 的一般形式

http://<主机>:<端口>/<路径>

冒号和两个斜线是规定的格式

使用 HTTP 的 URL

- 使用 HTTP 的 URL 的一般形式

http://<主机>:<端口>/<路径>

↑
这里写主机的域名

使用 HTTP 的 URL

- 使用 HTTP 的 URL 的一般形式

`http://<主机>:<端口>/<路径>`

↑
HTTP 的默认端口号是 80，通常可省略

使用 HTTP 的 URL

- 使用 HTTP 的 URL 的一般形式

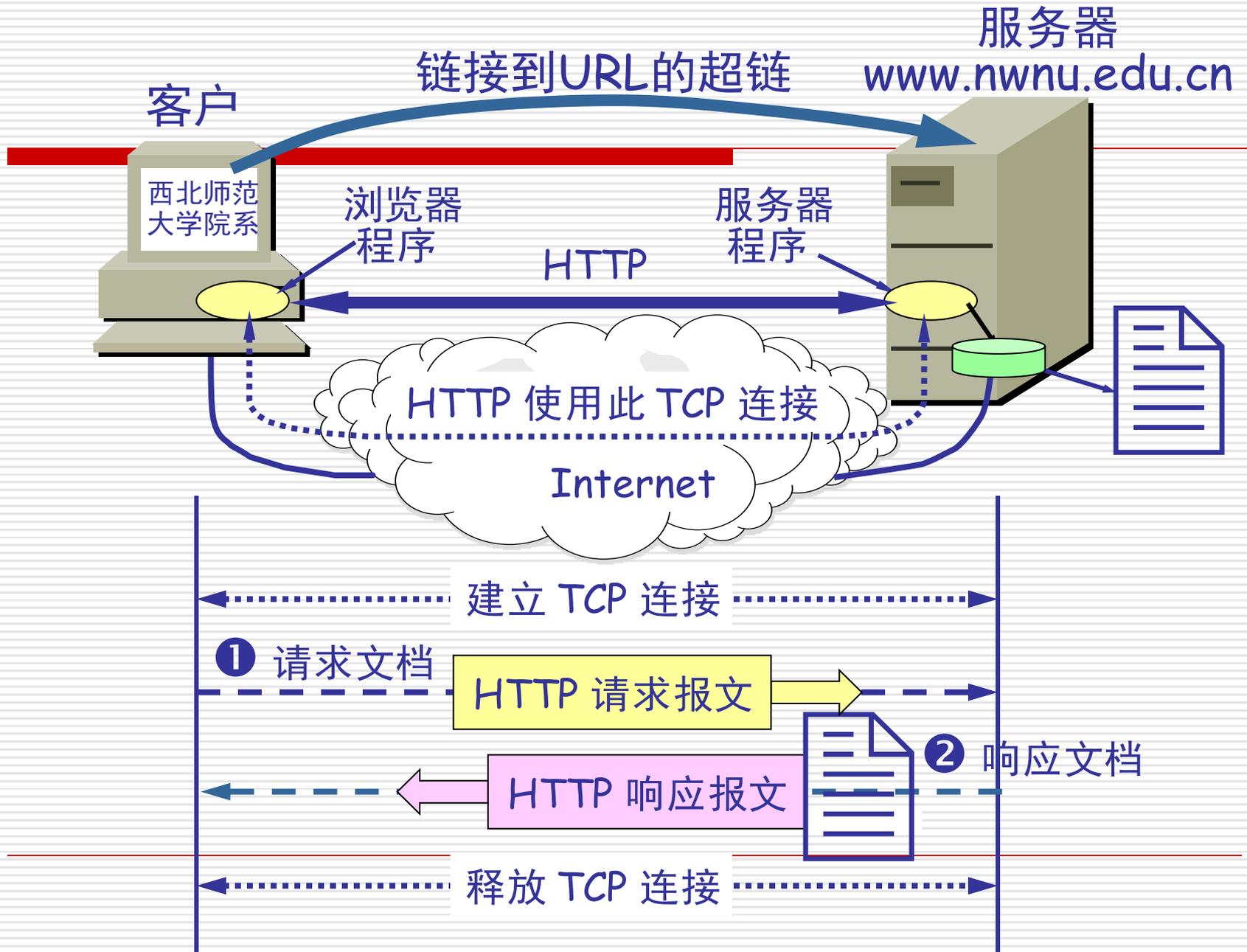
`http://<主机>:<端口>/<路径>`

若再省略文件的<路径>项，则 URL 就指到Internet上的某个**主页**(home page)。

3 超文本传送协议 HTTP

- 为了使超文本的链接能够高效率地完成，需要用 **HTTP** 协议来传送一切必须的信息。
 - 从层次的角度看，**HTTP** 是**面向事务的 (transaction-oriented)**应用层协议，它是万维网上能够可靠地交换文件（包括文本、声音、图像等各种多媒体文件）的重要基础。
-

万维网的工作过程



用户点击鼠标后所发生的事件 - 以西北师范大学为例

- (1) 浏览器分析超链指向页面的 **URL**。
 - (2) 浏览器向 **DNS** 请求解析 **www.nwnu.edu.cn** 的**IP** 地址。
 - (3) 域名系统 **DNS** 解析出西北师大服务器的 **IP** 地址。
 - (4) 浏览器与服务器建立 **TCP** 连接。
 - (5) 浏览器发出取文件命令。
 - (6) 服务器给出响应，把文件 **index.htm**(或其他主页文件)发给浏览器。
 - (7) **TCP** 连接释放。
 - (8) 浏览器显示“西北师范大学院系”文件 **index.htm** 中的所有文本。
-

4 万维网的文档

- 超文本标记语言 **HTML** 中的 **Markup** 的意思就是“设置标记”。
 - **HTML** 定义了许多用于排版的命令（即标签）。
 - **HTML** 把各种标签嵌入到万维网的页面中。这样就构成了所谓的 **HTML** 文档。**HTML** 文档是一种可以用任何文本编辑器创建的 **ASCII** 码文件。
-

HTML 文档

- 仅当 HTML 文档是以 `.html` 或 `.htm` 为后缀时，浏览器才对此文档的各种标签进行解释。
 - 如 HTML 文档改换以 `.txt` 为其后缀，则 HTML 解释程序就不对标签进行解释，而浏览器只能看见原来的文本文件。
 - 当浏览器从服务器读取 HTML 文档后，就按照 HTML 文档中的各种标签，根据浏览器所使用的显示器的尺寸和分辨率大小，重新进行排版并恢复出所读取的页面。
-

HTML 文档中标签的用法

<HTML>

HTML 文档开始

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

</HTML>

HTML 文档中标签的用法

<HTML>

<HEAD>

首部开始

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

</HTML>

HTML 文档中标签的用法

<HTML>

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

</HTML>

标题

HTML 文档中标签的用法

<HTML>

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

首部结束

<BODY>

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

</HTML>

HTML 文档中标签的用法

<HTML>

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

主体开始

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

</HTML>

HTML 文档中标签的用法

<HTML>

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

<H1>HTML 很容易掌握</H1>

1 级标题

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

</HTML>

HTML 文档中标签的用法

<HTML>

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

</HTML>

第一个段落

HTML 文档中标签的用法

<HTML>

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

</HTML>

第二个段落

HTML 文档中标签的用法

<HTML>

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

<P>这是第二个段落。</P>

</BODY>

主体结束

</HTML>

HTML 文档中标签的用法

<HTML>

<HEAD>

<TITLE>一个 HTML 的例子</TITLE>

</HEAD>

<BODY>

<H1>HTML 很容易掌握</H1>

<P>这是第一个段落。虽然很短，但它仍是一个段落。</P>

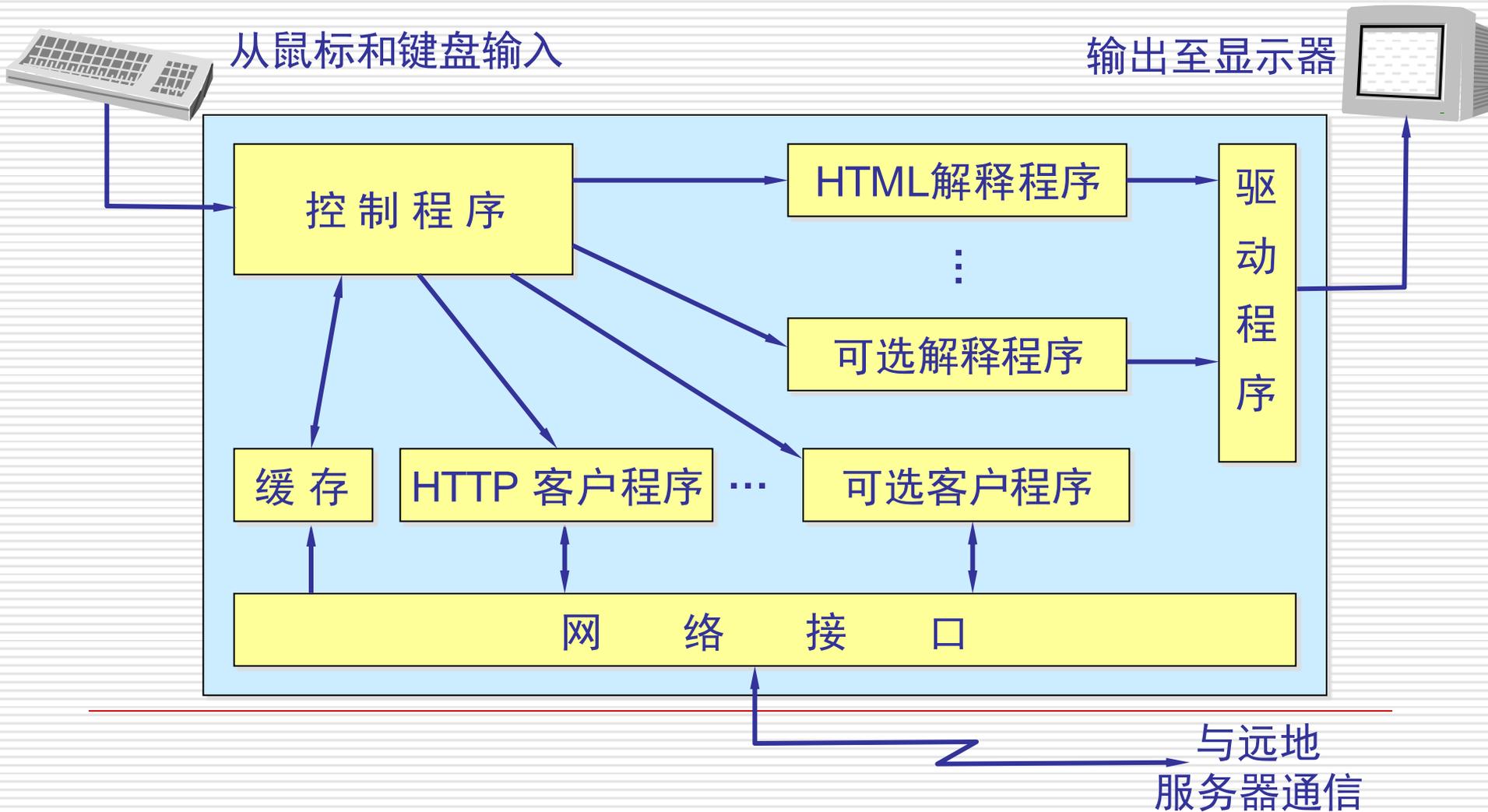
<P>这是第二个段落。</P>

</BODY>

</HTML>

HTML 文档结束

5 浏览器的结构



浏览器的主要组成部分

- 浏览器有一组客户、一组解释程序，以及管理这些客户和解释程序的控制程序。
 - 控制程序是其中的核心部件，它解释鼠标的点击和键盘的输入，并调用有关的组件来执行用户指定的操作。
-

解释程序

- **HTML 解释程序是必不可少的，而其他的解释程序则是可选的。**
 - **解释程序把 HTML 规格转换为适合用户显示硬件的命令来处理版面的细节。**
 - **许多浏览器还包含 FTP 客户程序，用来获取文件传送服务。**
 - **一些浏览器也包含电子邮件客户程序，使浏览器能够发送和接收电子邮件。**
-

浏览器中的缓存

- ❑ 浏览器将它取回的每一个页面副本都放入本地磁盘的缓存中。
 - ❑ 当用户用鼠标点击某个选项时，浏览器首先检查磁盘的缓存。若缓存中保存了该项，浏览器就直接从缓存中得到该项副本而不必从网络获取，这样就明显地改善浏览器的运行特性。
 - ❑ 但缓存要占用磁盘大量的空间，而浏览器性能的改善只有在用户再次查看缓存中的页面时才有帮助。
 - ❑ 许多浏览器允许用户调整缓存策略。
-

浏览器的功能

- 接收URL输入（超链点击）并发送请求报文
 - 显示并解释接收到的响应报文
 - 导航功能
 - 服务功能
 - 接受用户对浏览器的设置
-

浏览器的功能

- 浏览器可以用的协议（方法）包括：
 - **http://** （超文本传输协议）
 - **ftp://** （文件传输协议）
 - **gopher://** （菜单式检索）
 - **telnet://** （远程仿真终端协议）
 - **news:<新闻组>** （新闻组访问）
 - **mailto:<邮件地址>** （邮件访问）
-

浏览器的功能

- 浏览器屏幕可显示的内容包括
 - 网页正文
 - 超链信息，导航按钮（引导到其他网页）
 - 图像信息（一般为jpg格式的压缩图像）
 - 音频/视频信息（音乐、电影）
 - 动态信息（动态的广告信息）
 - 进行交互功能的表单（如电子邮件用户的注册表格）
-

浏览器的功能

□ 导航功能

- 向前，向后（翻到曾经显示的前/后一个网页）
 - 打开，搜索（将文件系统中存储的网页文件打开）
 - 历史资料（根据时间检索曾经访问的网页）
 - 插入书签（将重要的网页访问连接存入系统）
-

浏览器的功能

□ 服务功能

- 打印（将网页从打印机输出）
 - 邮件发送（将网页作为邮件发送）
 - 存成文件（将网页和所携信息分别存储）
 - 查看网页代码（查看代码，学习网页设计）
 - 帮助信息(理解浏览器的应用)
 - 刷新（请求最新的网页内容）
 - 停止（中断网页的下载过程）
-

8.2 Internet基本服务与应用

8.2.2 FTP

- **文件传送协议 FTP (File Transfer Protocol)** 是Internet上使用得最广泛的文件传送协议。
 - **FTP** 提供交互式的访问，允许客户指明文件的类型与格式，并允许文件具有存取权限。
 - **FTP** 屏蔽了各计算机系统的细节，因而适合于在异构网络中任意计算机之间传送文件。
-

8.2.2 文件传送协议

网络环境下复制文件的复杂性：

- (1) 计算机存储数据的格式不同。
 - (2) 文件的目录结构和文件命名的规定不同。
 - (3) 对于相同的文件存取功能，操作系统使用的命令不同。
 - (4) 访问控制方法不同。
-

8.2.2 文件传送协议

- ◆ FTP可以实现上传和下载两种文件传输方式
- ◆ FTP提供了匿名(**anonymous**)访问和授权访问两种访问方式
- ◆ FTP可以传输几乎所有类型的文件
- ◆ 利用**FTP**传输文件有三种方式:

FTP命令行、浏览器、FTP专用工具

8.2.3 按远程终端协议 TELNET

- **TELNET** 是一个简单的远程终端协议，也是 **Internet** 的正式标准。
 - 用户用 **TELNET** 就可在其所在地通过 **TCP** 连接注册（即登录）到远地的另一个主机上（使用主机名或 **IP** 地址）。
 - **TELNET** 能将用户的击键传到远地主机，同时也能将远地主机的输出通过 **TCP** 连接返回到用户屏幕。这种服务是透明的，因为用户感觉到好像键盘和显示器是直接连在远地主机上。
-

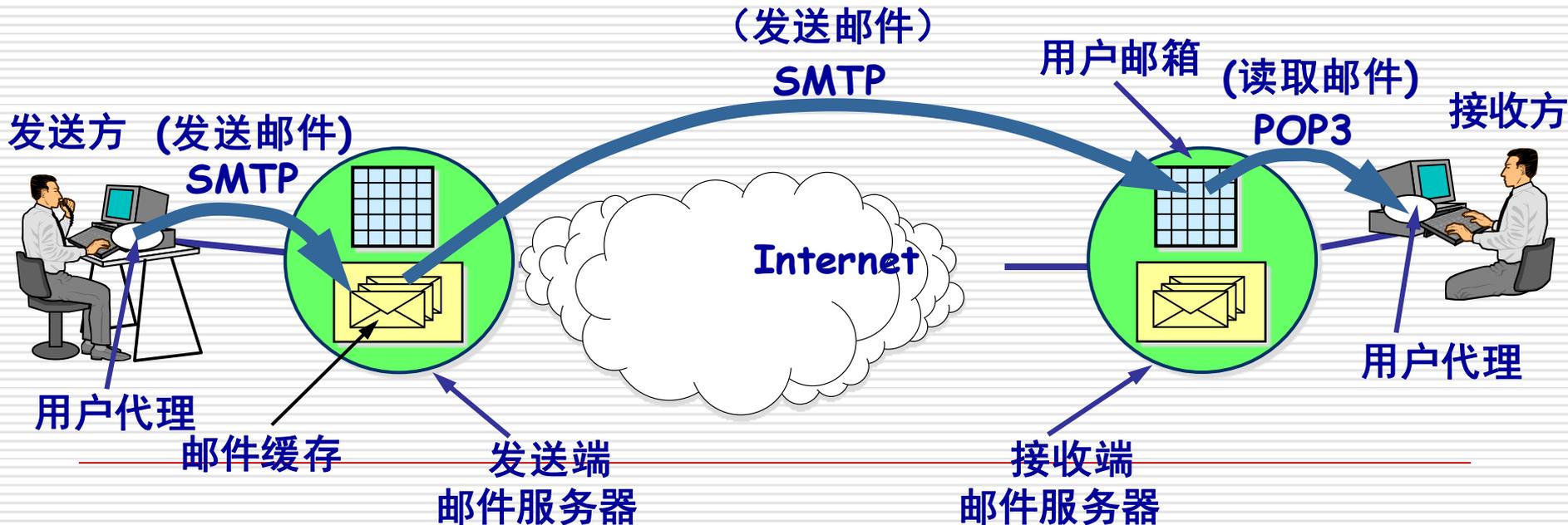
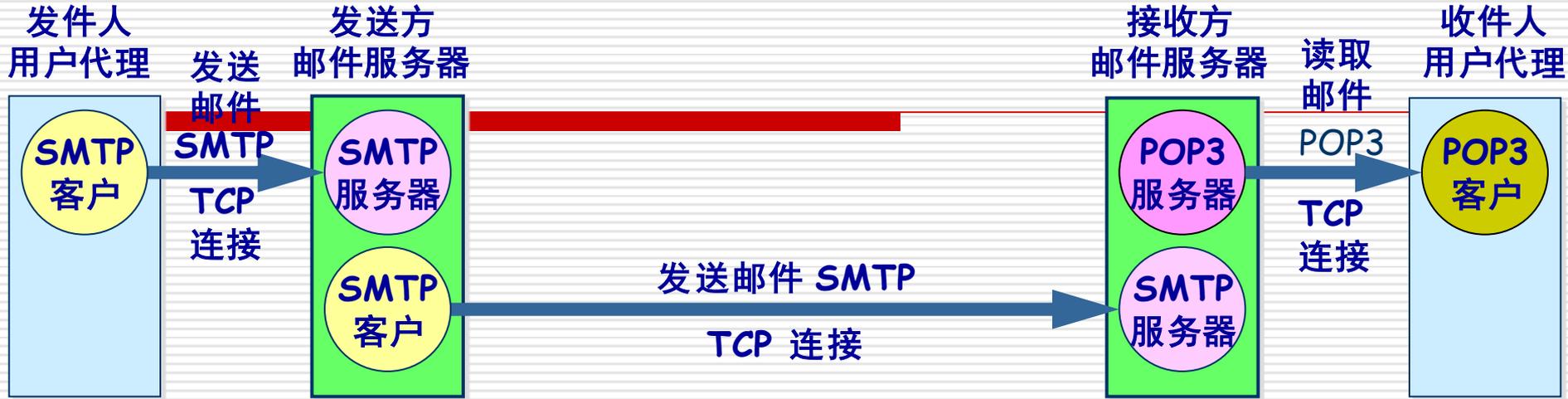
8.2.4 电子邮件

- ❑ 电子邮件(e-mail)是Internet上使用得最多的和最受欢迎的一种应用。
- ❑ 电子邮件把邮件发送到收件人使用的邮件服务器，并放在其中的收件人邮箱中，收件人可随时上网到自己使用的邮件服务器进行读取。
- ❑ 电子邮件不仅使用方便，而且还具有传递迅速和费用低廉的优点。
- ❑ 现在电子邮件不仅可传送文字信息，而且还可附上声音和图像。

电子邮件的一些标准

- 发送邮件的协议：**SMTP**和**MIME**
 - 读取邮件的协议：**POP3** 和 **IMAP4**
 - **MIME** 在其邮件首部中说明了邮件的数据类型(如文本、声音、图像、视像等)，使用 **MIME** 可在邮件中同时传送多种类型的数据。
-

电子邮件的最主要的组成构件



用户代理 UA (User Agent)

- 用户代理 **UA** 就是用户与电子邮件系统的接口，是电子邮件客户端软件。
 - 用户代理的功能：撰写、显示、处理和通信。
 - 邮件服务器的功能是发送和接收邮件，同时还要向发信人报告邮件传送的情况（已交付、被拒绝、丢失等）。
 - 邮件服务器按照客户服务器方式工作。邮件服务器需要使用发送和读取两个不同的协议。
-

发送和接收电子邮件的几个重要步骤

- ❶ 发件人调用 **PC** 机中的用户代理撰写和编辑要发送的邮件。
 - ❷ 发件人的用户代理把邮件用 **SMTP** 协议发给发送方邮件服务器，
 - ❸ **SMTP** 服务器把邮件临时存放在邮件缓存队列中，等待发送。
 - ❹ 发送方邮件服务器的 **SMTP** 客户与接收方邮件服务器的 **SMTP** 服务器建立 **TCP** 连接，然后就把邮件缓存队列中的邮件依次发送出去。
-

发送和接收电子邮件的几个重要步骤（续）

- ⑤ 运行在接收方邮件服务器中的**SMTP**服务器进程收到邮件后，把邮件放入收件人的用户邮箱中，等待收件人进行读取。
 - ⑥ 收件人在打算收信时，就运行 **PC** 机中的用户代理，使用 **POP3**（或 **IMAP4**）协议读取发送给自己的邮件。
 - 请注意，**POP3** 服务器和 **POP3** 客户之间的通信是由 **POP3** 客户发起的。
-

电子邮件的组成

- 电子邮件由信封(envelope)和内容(content)两部分组成。
 - 电子邮件的传输程序根据邮件信封上的信息来传送邮件。用户在从自己的邮箱中读取邮件时才能见到邮件的内容。
 - 在邮件的信封上，最重要的就是收件人的地址。
-

电子邮件地址的格式

- TCP/IP 体系的电子邮件系统规定电子邮件地址的格式如下：

收件人邮箱名@邮箱所在主机的域名

- 符号“@”读作“at”，表示“在”的意思。
- 例如，电子邮件地址 goupz@163.com

这个用户名在该域名的范围内是唯一的。

邮箱所在的主机的域名在全世界必须是唯一的

8.2.5 网络交流

□ 1. 新闻组Usenet

- Usenet并不是一个网络系统，只是建立在Internet上的逻辑组织，不同于Internet上的交互式操作方式。
 - 在Usenet服务器上存储的各种信息，会周期性地转发给其他Usenet服务器，Usenet的基本通信方式是电子邮件，但它不是采用点对点通信方式，而是采用多对多的传递方式。
 - Usenet 包括各种主题的论坛，每一个主题就是一个新闻组，这些新闻组覆盖了从科研、教育到新闻、体育、文化、宗教等方面，几乎包括人类关心的所有话题。
-

8.2.5 网络交流

□ 2. 电子公告牌BBS

- 电子公告牌**BBS** (**Bulletin Board System**) 是**Internet**上较常用的服务功能之一，最早的**BBS**用来公布股市价格等信息，采用基于远程登录的服务，想使用**BBS**服务的用户，必须首先利用远程登录功能登录到**BBS**服务器，**BBS**服务器通过计算机远程访问把各类共享信息、资源以及联系提供给各类用户。
- **BBS**按照不同的主题、分主题分为很多布告栏。
- 清华大学的水木清华<http://bbs.tsinghua.edu.cn>
- 北京大学的北大未名<https://bdwm.net>
- 复旦大学的日月光华<http://bbs.fudan.edu.cn>
- 人民大学的网上人大<http://www.cmr.com.cn>
- 电子科技大学的一网深情<http://bbs.uestc.edu.cn>

8.2.5 网络交流

□ 3. QQ与MSN

- MSN与QQ相比最大的特点是简洁和保密，使用**MSN**可以只与想交流的人进行沟通，即“不和陌生人说话”，对话内容不会被公开，发送文件非常方便，不会受到防火墙的限制。而**QQ**发送文件时往往要求双方用户必须在同一个防火墙内。
-

8.2.5 网络交流

□ 4. 网络博客

- 与BBS相比较，**Blog**更强调的是个性，是利用现成的网页模板，发表文章和评论，并可按日历索引的个人网站。
-

8.2.6 电子商务

- ◆ 电子商务（**EC, Electronic Commerce**）是利用**Internet**进行商务活动的方式，是**21**世纪人类信息世界的核心，具有高效率、价格低廉、高收益、全球化等优点。
 - ◆ 电子商务的处理方式有**B2B**、**B2C**、**C2C**三种。
-

8.2.6 电子商务

- (1) **B2B (Business to Business, 企业对企业)**：是企业之间通过**Internet**进行产品、服务、贸易的交换。如阿里巴巴 (**China.alibaba.com**)。
 - (2) **B2C (Business to Consumer, 企业对消费者)**：是企业与消费者之间通过**Internet**进行的销售产品和服务方式。例如，最具代表的是网上商业零售网站亚马逊网上商店、当当网**www.dangdang.com**、卓越网**www.joyo.com**等。
 - (3) **C2C (Consumer to Consumer, 消费者对消费者)**：是消费者之间通过**Internet**提供销售产品和服务的方式。此类网站最具代表的是淘宝网**www.taobao.com**。
-

8.3 网络信息检索技术

8.3.1 网络信息检索及其分类

- 网络信息检索是将网络信息按一定方式存储起来，用科学的方法，利用检索工具，为用户检索、揭示、传递知识和信息的业务过程。
 - 按检索内容分类：数据信息检索、事实信息检索和文献信息检索
 - 按照组织方式的不同，有全文检索、超文本检索和超媒体检索
-

8.3 网络信息检索技术

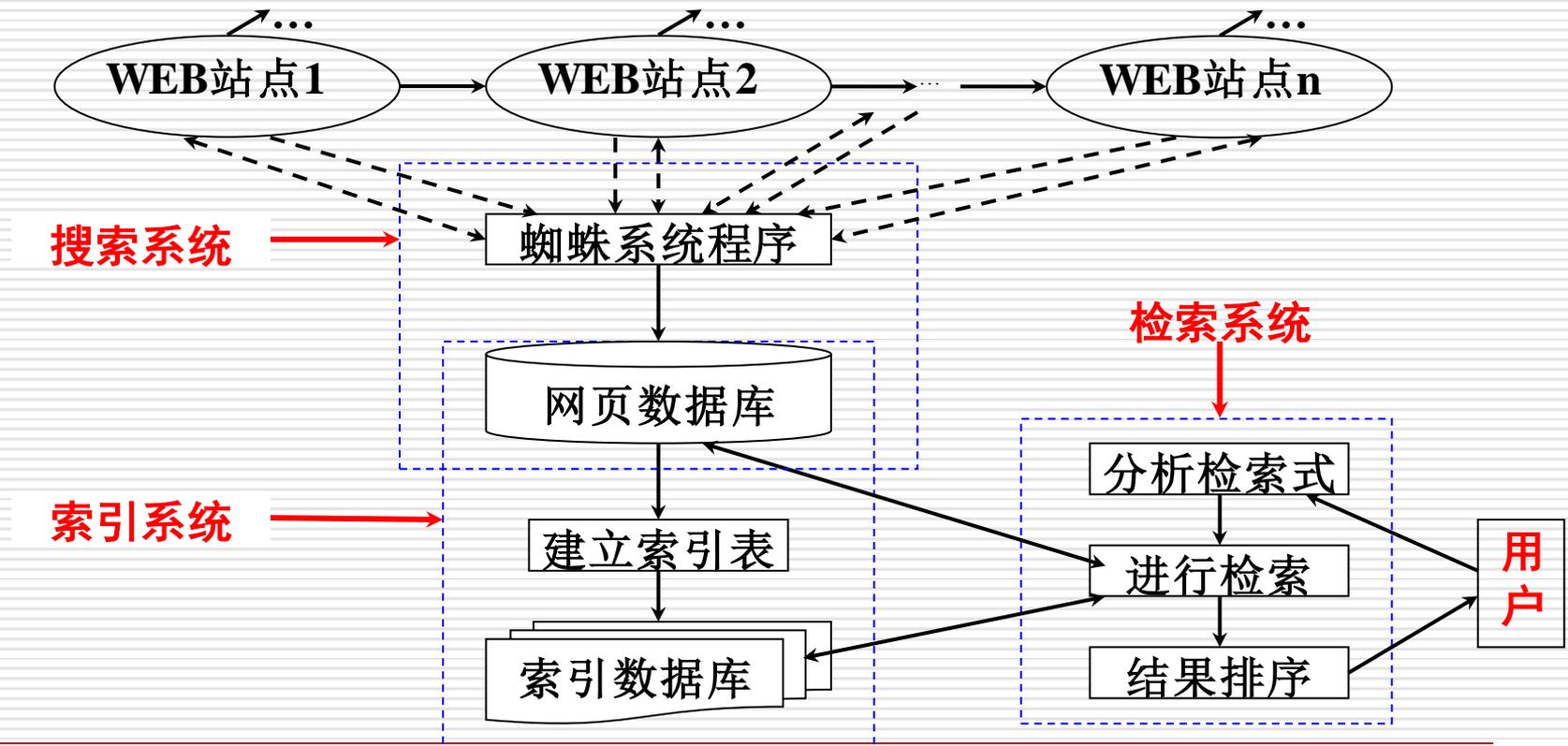
8.3.2 网络信息检索技术

网络信息资源检索工具主要有：

- ① 基于菜单式的检索工具和基于关键词的检索工具，如**Gopher**、**WAIS**等。
 - ② 基于超文本的检索工具，如**WWW**方式检索。
 - ③ 网络指南工具搜索引擎。
 - ④ 全文数据库检索工具和多媒体信息检索工具。
-

8.3.3 搜索引擎

1. 搜索引擎概念与原理



搜索引擎工作原理

1 搜索引擎概念与原理

搜索引擎由**4**部分组成：

搜索系统、索引系统、检索系统、用户接口

搜索引擎工作原理可以概括为：“蜘蛛”（**Spider**）系统程序抓取网页→索引系统建立索引数据库→检索系统在索引数据库中搜索排序→页面生成系统将结果返回给用户。

2 搜索引擎分类

从搜索引擎的工作原理和组织形式划分

- ◆ 全文搜索引擎
 - ◆ 目录索引类搜索引擎
 - ◆ 元搜索引擎
-

1)全文搜索引擎(Full Text Search Engine)

- 检索程序以某种策略自动在互联网中搜集发现信息
 - 索引系统为搜集到的信息建立索引数据库
 - 检索系统根据用户的查询输入检索索引库
 - 如果找到与用户要求内容相符的网站，采用特殊的算法计算出各网页的信息关联程度，根据关联程度高低，按顺序将这些网页链接制成索引返回给用户
 - ◆ 优点是信息量大、更新及时；缺点是返回信息过多，有很多无关信息，用户必须从结果中进行筛选。
 - ◆ 代表性的有**Google**、百度等。
-

2) 目录搜索引擎(Search Index/Directory)

- ◆ 目录搜索引擎只是一些按照目录分类的网站超链接列表，主要通过人工发现信息，并依靠标引人员对信息进行分析和分类，由专业人员手工建立关键词索引，建立目录分类体系。
 - ◆ 最具代表性的是**Yahoo!**、新浪等。
-

3) 元搜索引擎 (Meta Search Engine)

调用其他独立搜索引擎的引擎。

在接收用户查询请求后，同时向多个搜索引擎递交，将返回的结果进行重复排除、重新排序等处理后，作为自己的结果返回给用户。

◆ 搜索效果始终不太理想，没有一个特别具有优势的元搜索引擎。

3 搜索引擎的使用

一般的搜索引擎都具有“分类检索”和“关键词检索”两种搜索方法。

1) 分类检索

□ “分类检索”是在搜索网站中，按照主题分类，在相关的类别中进行查询，操作比较简单。

2) 关键词检索

□ “关键词检索”是利用标题、词语等关键字，从搜索引擎数据库中准确地查找相关的信息。

2) 关键词检索

(1) 使用布尔 (Boolean) 检索

- ◆主要布尔逻辑关系词有“与”、“或”、“非”等。
- ◆搜索引擎基本都支持布尔逻辑命令查询，用好这些命令符号可以大幅提高搜索精度。

① 逻辑“与”。**“and”**或**“+”**，只有含“与”的关键词全部出现时，搜索到的结果才符合条件。**Google**中不使用**“and”**，但可在检索词前加上空格**“+”**表示必须包含该词条。

② 逻辑“或”。**“or”**，只要“或”的关键词中有任何一个出现，搜索到的结果就算符合条件。

③ 逻辑“非”。**“not”**或**“-”**，搜索结果不应含有“非”后面的关键词。

汉字作关键词时，不要随意加空格，因为许多搜索引擎把空格认作特殊操作符，有的与**“and”**一样，有的与**“or”**一样。

2) 关键词检索

(2) 精确搜索的应用

- ◆ 精确搜索符引号（“ ”）括起来的词表示要精确匹配，即将关键词或关键词的组合作为一个字符串在其数据库中进行搜索，不包括演变形式。
 - ◆ 书名号（“《》”）为百度独有的，加上书名号的查询词，有两层特殊功能，一是书名号会出现在搜索结果中；二是被书名号扩起来的内容，不会被拆分。
-

2) 关键词检索

(3) 通配符 “*” 或 “?” 的使用

大多数搜索引擎将 “*” 和 “?” 作为通配符使用，“*” 可以代替任意几个字符，“?” 代表单个字符。

2) 关键词检索

(4) 字段检索

网络信息实际上不分字段，但有的搜索引擎设计了类似于字段检索的功能，运用字段设置，可以把检索词限制在一定位置范围内。

(4) 字段检索

① 检索结果限定在网页标题中

◆把查询范围限定在网页标题中，有时能获得良好的效果。

◆一般使用“**intitle:检索词**”

(4) 字段检索

② 检索结果限定在特定网站中

◆ 如果知道某个网站或者域名中有需要搜索的内容，可以使用“检索词 **site:** 站点域名”（不含“**http://**”及“**/**”等），把结果限制在某个网站或者域名之内。

◆ 如果要排除某网站或者域名范围内的页面，只需用“-网站/域名”。

(4) 字段检索

③ 将搜索范围限定在**URL**链接中

◆对搜索结果的**URL**做某种限定，可以获得良好的效果。

◆**Google**中的“**allinurl:检索词**”，返回的网页链接中包含所有查询关键字。这个查询的对象只集中于网页的链接字符串。

(4) 字段检索

③ 将搜索范围限定在URL链接中

◆ **related**用来搜索结构内容方面相似的网页，如搜索所有与新浪主页相似的页面（如网易首页，搜狐首页等），使用搜索“**related:www.sina.com.cn/index.shtml**”即可。

◆ **Google**和**百度**中，“**inurl:检索词**”返回的网页链接中包含第一个关键字，后面的关键字则出现在链接中或者网页文档中。有很多网站把某一类具有相同属性的资源名称显示在目录名称或者网页名称中，如“**MP3**”等，因此，可以用**inurl**找到这些相关资源链接，用第二个关键词确定是否有某项具体资料。**inurl**通常能够提供非常精确的专题资料。

(5) 找不同类型的信息

- ◆在Internet上很多有价值的资料并非是普通的网页，而是以pdf、doc、xls、ppt、rtf、swf等文档类型格式存在的。
 - ◆百度支持对Office文档（Word、Excel、Powerpoint）、pdf文档、rtf文档进行全文搜索。
 - ◆Google支持13种非HTML文件的搜索。
 - ◆搜索这类文档，只需要在普通查询词后面加“filetype:文档类型”限定。
 - ◆“Filetype:”后可跟以下文件格式：pdf、doc、xls、ppt、rtf、swf、all等。
 - ◆其中，all表示搜索所有这些文件类型。
-

8.3.4 数字图书馆

1) 超星电子图书馆

登录超星数字图书馆或学校的超星数字图书馆或其它下载网站，均可以下载SSReader（超星）阅览器（超星数字图书是按页组成的PDG文件，阅览器阅读需要下载并安装超星阅览器，而IE阅读时自动下载IE阅读插件）。超星数字图书馆提供了三种阅读方式：

- ① 免费阅览室阅读：进入免费阅览室→查找所需图书；
 - ② 会员图书馆阅读：进入会员图书馆→订阅会员服务→查找所需图书；
 - ③ 电子书店阅读：进入电子书店 →查找所需图书→付费购买成功。查找到图书后，单击“阅览器阅读”或“IE阅读”按钮浏览图书。
-

8.3.4 数字图书馆

2) 中国期刊网 (CNKI)

- (1) 简单检索
- (2) 标准检索
- (3) 高级检索
- (4) 专业检索
- (5) 检索结果查看与下载
- (6) 全文浏览

CAJViewer浏览器是中国知网的专用全文格式阅读器，支持CAJ、NH、KDH 和 PDF格式文件，可以在线阅读中国知网的原文，也可以阅读下载到本地硬盘的CNKI文献，打印效果可以达到与原版显示一致的程度。

8.4 网络信息安全技术

8.4.1 网络信息安全技术概述

1. 网络信息安全的概念

- ◆信息安全是信息系统资源不受自然和人为有害因素的威胁和危害。
 - ◆解决信息安全问题应该同时从技术和管理两方面着手。
-

2. 网络系统安全威胁

对网络系统的威胁有很多方面，主要是人为攻击和自然灾害攻击。

1) 被动攻击

在不影响网络正常工作的情况下，进行截获、窃取、破译以获得重要机密信息。主要有搭线监听和无线截获攻击。

2) 主动攻击

以各种方式有选择地破坏信息的有效性、完整性。包括：①假冒、②重放、③入侵、④篡改、⑤拒绝服务、⑥抵赖、⑦服务欺骗、⑧阻塞、⑨病毒等。

8.4.1 网络信息安全技术概述

3. 网络安全策略

1) 物理安全策略

2) 访问控制策略

① 入网访问控制

② 网络的权限控制

③ 目录级安全控制

④ 属性安全控制

⑤ 网络服务器安全控制

⑥ 网络监测和锁定控制

⑦ 网络端口和节点的安全控制

8.4.2 计算机病毒防范技术

计算机病毒：

指编制或者在计算机程序中插入破坏计算机功能或者毁坏数据，影响计算机使用，并能自我复制的程序代码。

计算机病毒的特征：

人为的特制程序，具有自我复制能力，很强的传染性，一定的潜伏性，特定的触发性，很大的破坏性和不可预见性。

计算机病毒的分类：

- (1) 引导型病毒
- (2) 文件型病毒
- (3) 宏病毒
- (4) 蠕虫病毒
- (5) 脚本病毒
- (6) 复合型病毒
- (7) 木马程序

8.4.3 网络安全技术

1. 防火墙技术

防火墙（Firewall）是用于防止外部网络的恶意攻击对内部网络造成不良影响而设置的安全防护措施。

2. 加密技术

1) 对称加密技术

在加密和解密过程中都必须用到同一个秘密密钥的加密体制。

2) 非对称加密技术

也称公钥密码加密技术。它的加密密钥和解密密钥是两个不同的密钥，一个称为公开密钥也称公钥，另一个称为私有密钥也称私钥。

3) 数字签名技术

一般采用非对称加密技术，以电子形式存储的一种消息，可以在网络中传输。其安全性取决于所采用的密码体制的安全程度。

8.4.3 网络安全技术

3. 数字证书与数字认证

1) 数字证书 即数字ID，是网络中标志通信各方身份信息的一系列数据，提供了一种在Internet上验证身份的方式，方便地保证由鲜为人知的网络发来信息的可靠性，同时建立收到信息的拥有权及完整性。主要应用在电子政务、网上购物、安全电子邮件、网上交易、网上银行、企业与企业的电子贸易等方面。

2) 数字认证 是检查一份给定的证书是否可用的过程，也称证书验证。数字认证引入了一种机制来确保证书的完整性和证书颁发者的可信赖性。

8.4.3 网络安全技术

4. 检测技术

- 1) 入侵检测
- 2) 漏洞扫描
- 3) 网络黑客

5. 其他安全技术

- 1) IC卡技术
 - 2) 面像识别技术
-