



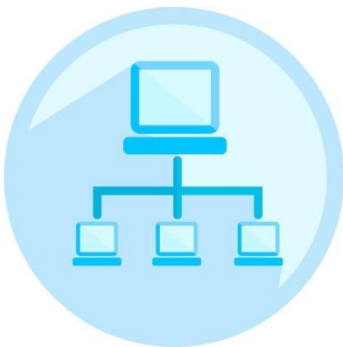
计算机网络



顾 军

计算机学院

jgu@cumt.edu.cn





专题6：互联网提供了哪些高层应用

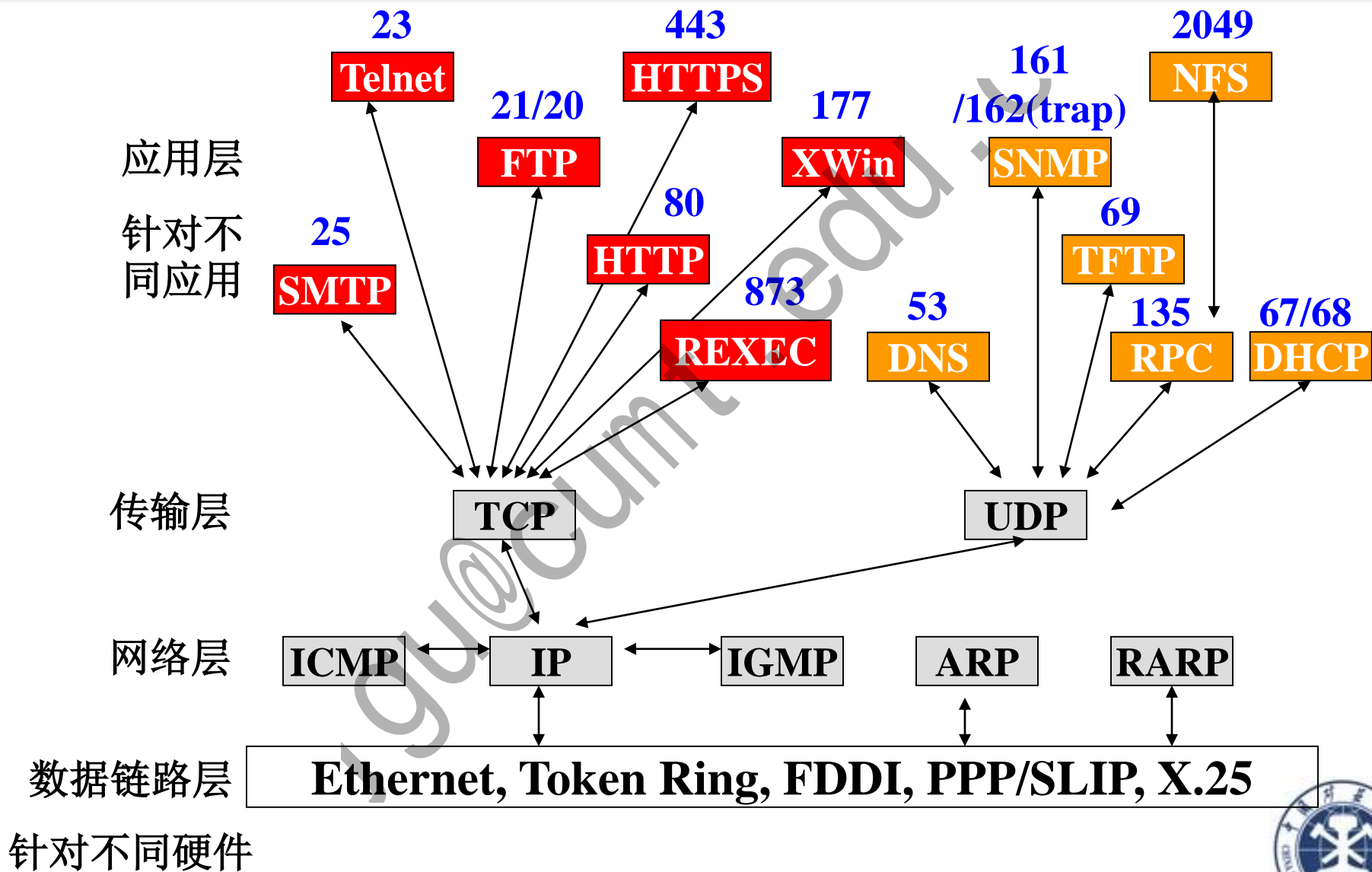


- 应用层(application layer)
- 运输层(transport layer)
- 网络层(network layer)
- 数据链路层(data link layer)
- 物理层(physical layer)





TCP/IP协议族中的应用层协议





Q11: Web网站是怎么回事?

- 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因特网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**”。

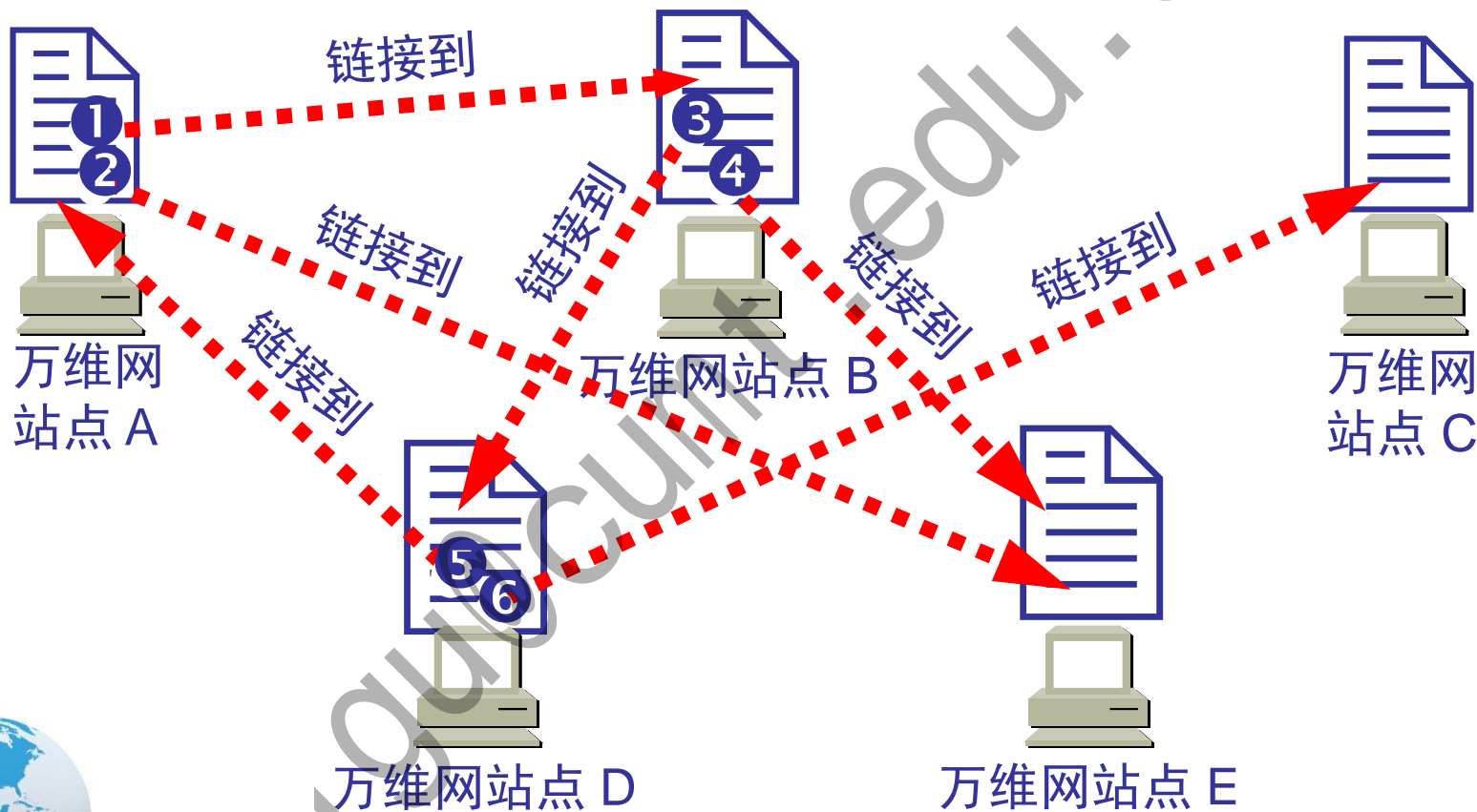


蒂姆·伯纳斯-李，英国计算机科学家，万维网的发明者，麻省理工学院教授。他的父母都参与了世界上第一台商业电脑，曼切斯特1型（Manchester Mark I）的建造。1990年12月25日，罗伯特·卡里奥在CERN和他一起成功通过Internet实现了HTTP代理与服务器的第一次通讯。





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，它是**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因特网上的超文本系统中。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。





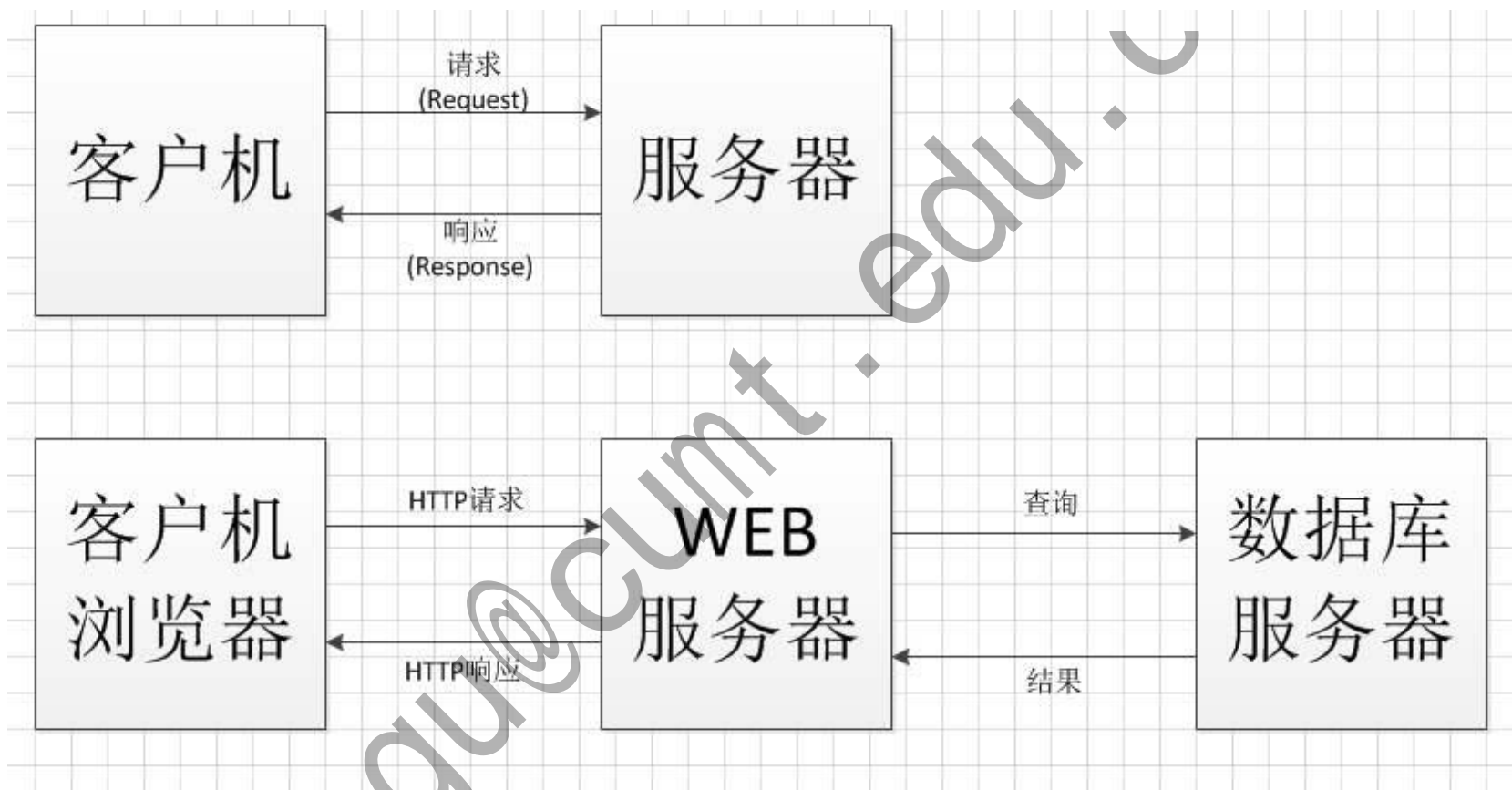
Q12: 万维网的工作方式 ?

- 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)。





万维网工作架构





多层B/S工作模式





Q13: 怎么标识万维网文档 ?

- 统一资源定位符 URL (Uniform Resource Locator)使每一个文档在整个因特网的范围内具有唯一的标识符 URL。
- URL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。

<协议>://<主机>[:<端口>]/<路径>





URL 的一般形式

<协议>://<主机>:<端口>/<路径>

默认端口号是 80
有时可省略

<主机> 是存放资源的主机
在因特网中的域名

ftp —— 文件传送协议 FTP

http —— 超文本传送协议 HTTP

News —— USENET 新闻





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 就指到因特网上的某个主页(home page)。





Q14: 如何编写万维网文档?

- 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因特网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。





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 的格式与标签

- **元素**(element)是 HTML 文档结构的基本组成部分。一个 HTML 文档本身就是一个元素。每个 HTML 文档由两个主要元素组成：**首部**(head)和**主体**(body)。
- 首部包含文档的**标题**(title)，以及系统用来标识文档的一些其他信息。标题相当于文件名。
- 文档的**主体**是 HTML 文档的最主要的部分。
- 主体部分往往又由若干个更小的元素组成，如**段落**(paragraph)、**表格**(table)、和**列表**(list)等。





HTML 的标签

- HTML用一对标签（即一个开始标签和一个结束标签）或几对标签来标识一个元素。
- 开始标签由一个小于字符“<”、一个标签名、和一个大于字符“>”组成。
- 结束标签和开始标签的区别只是在小于字符的后面要加上一个斜杠字符“/”。
- 标签名并不区分大写和小写。
- 有一些标签可以将结束标签省略。





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>





插入图像

开始标签

结束标签

插入图像

高度是 100 像素

宽度是 65 像素

插入的图像文件名是 portrait.gif





万维网页面中的超链

- 定义一个超链的标签是 `<A>`。字符 A 表示 **锚** (Anchor)。
- 在HTML文档中定义一个超链的语法是：

`X`

超链的起点

这个地方填写超链终点的 URL





链接举例

中国矿业大学

超链的终点是
中国矿业大学的
主页

超链的起点是
某个文档中的
这六个字





链接方式

- 本地链接：超链指向本计算机中的某个文件。
- 远程链接：超链的终点是其他网点上的页面。





本地链接

本地链接可进行许多的简化：

- 协议(`http://`)被省略——表明与当前页面的协议相同。
- 主机域名被省略——表明是当前的主机域名。
- 目录路径被省略——表明是当前目录（对于远程链接，表明是主机的默认根目录）。
- 文件名被省略——表明是当前文件（对于远程链接，表明是对方服务器上默认的文件名，通常是一个名为 `index.html` 的文件）。





相对路径名与绝对路径名

- 使用简化的 URL，在 HREF= 的后面使用的是相对路径名。
- 使用完整的 URL 则是使用绝对路径名。
- 使用相对路径名的好处不仅是可以减少键入一些字符，而且也便于目录的改动。





本文件链接举例





远程链接举例

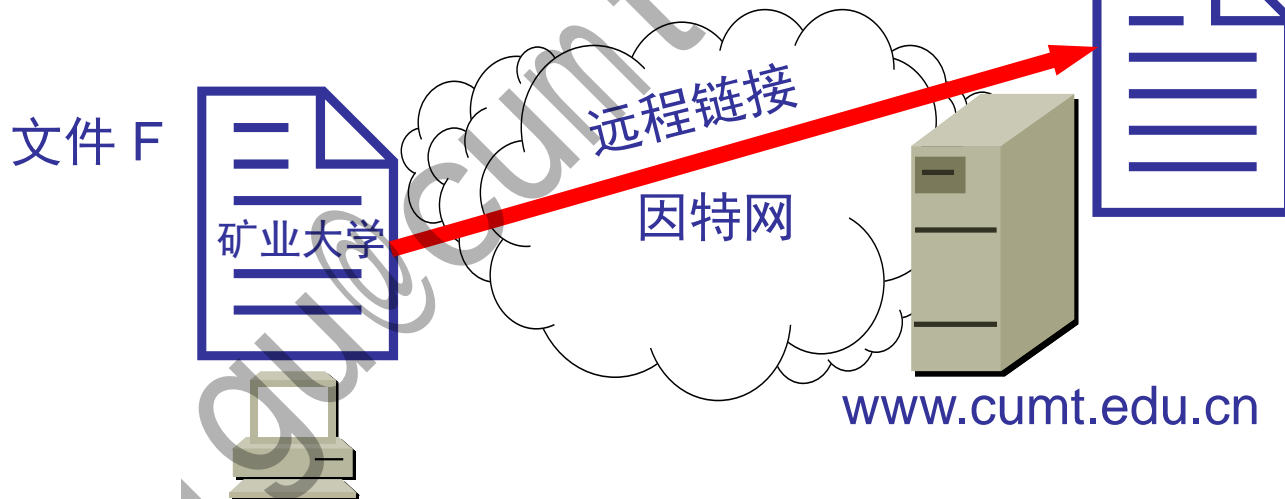
远程链接：

 中国矿业大学

终点

起点

中国矿业大学
主页





CSS

- CSS (Cascading Style Sheets) 是层叠样式表，它是一种样式表语言，用于为 HTML 文档定义布局。
- CSS 与 HTML 的区别就是：HTML 用于结构化内容，而 CSS 则用于格式化结构化的内容。





XML

- XML (Extensible Markup Language)是可扩展标记语言，它和HTML很相似。
- XML的设计宗旨是传输和存储数据，焦点是数据内容，是独立于软件和硬件的信息传输工具。
- HTML是用来显示数据（HTML是为了在浏览器上显示数据），焦点是数据的外观。
- XML不是要替换HTML，而是对HTML的补充。





XHTML

- XHTML (Extensible HTML) 是可扩展超文本标记语言，它与 HTML 4.01 几乎是相同的。
- 但 XHTML 是更严格的 HTML 版本，也是一个 W3C 标准（2000年1月），是作为一种 XML 应用被重新定义的 HTML，并将逐渐取代 HTML。
- 新的浏览器都支持 XHTML。





HTML5



- 2014年10月29日，万维网联盟宣布，经过接近8年的艰苦努力，HTML5标准规范终于制定完成。
- HTML5的设计目的是为了在移动设备上支持多媒体。
- HTML5将会取代1999年制定的HTML 4.01、XHTML 1.0标准，以期能在互联网应用迅速发展的时候，使网络标准达到符合当代的网络需求，为桌面和移动平台带来无缝衔接的丰富内容。





Q15: 怎么访问万维网资源 ?

- 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，使用TCP连接进行可靠传送。





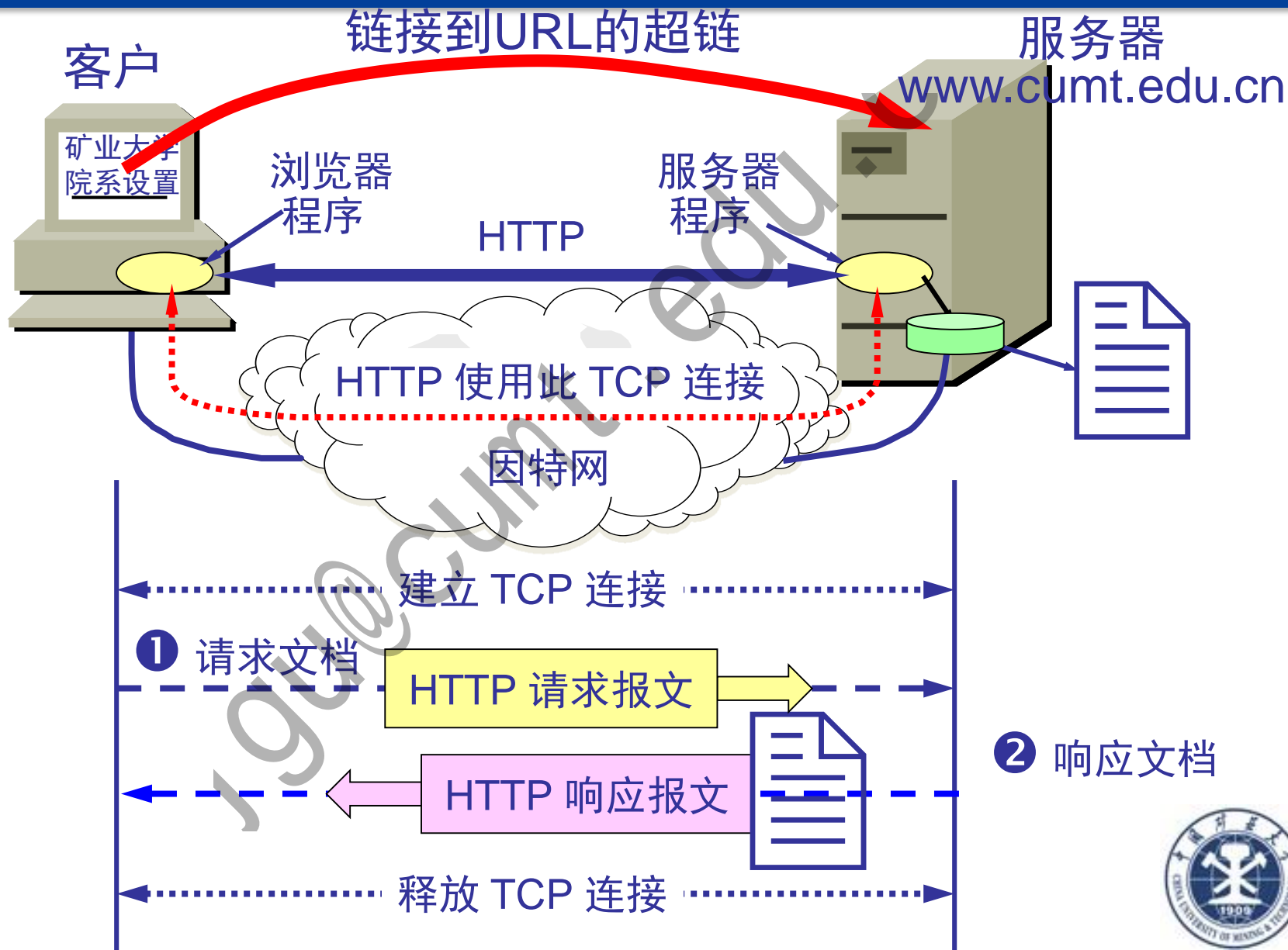
HTTP 的操作过程

- 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。
 - 事务，一般是指要做的或所做的事情。在计算机术语中是指访问并可能更新数据库中各种数据项的一个程序执行单元(unit)。
 - HTTP事务=请求命令+响应结果（通过HTTP报文的格式化数据块进行的）。





万维网的工作过程





用户点击鼠标后所发生的事件

- (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.cumt.edu.cn` 的 IP 地址。
- (3) 域名系统 DNS 解析出中国矿业大学服务器的 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：
`GET /chn/yxsx/index.htm`。
- (6) 服务器给出响应，把文件 `index.htm` 发给浏览器。
- (7) TCP 连接释放。
- (8) 浏览器显示“矿业大学院系设置”文件 `index.htm` 中的所有文本。





HTTP的不同版本

版本	产生时间	内容	发展现状
HTTP/0.9	1991年	不涉及数据包传输，规定客户端和服务端之间通信格式，只能GET请求	没有作为正式的标准
HTTP/1.0	1996年	传输内容格式不限制，增加PUT、PATCH、HEAD、OPTIONS、DELETE命令	正式作为标准，RFC 1945
HTTP/1.1	1997年	持久连接(长连接)、节约带宽、HOST域、管道机制、分块传输编码	2015年前使用最广泛，RFC 2616
HTTP/2	2015年	多路复用、服务器推送、头信息压缩、二进制协议等	逐渐覆盖市场





Q16: HTTP/1.0如何工作?

- HTTP 协议本身是**无连接**的，虽然它使用了面向连接的 TCP 向上提供的服务。
 - 通信的双方在交换HTTP之前不需要先建立HTTP连接。
 - 无连接：服务器处理完客户的请求，并收到客户的应答后，即断开连接。
- HTTP/ 1.0 协议是**无状态**的(stateless)。
 - 无状态是指服务器不知道客户端是什么状态。
 - 每次的请求都是独立的，它的执行情况与结果与前面的请求和之后的请求是无直接关系的，它不会受前面的请求应答情况直接影响，也不会直接影响后面的请求应答情况。





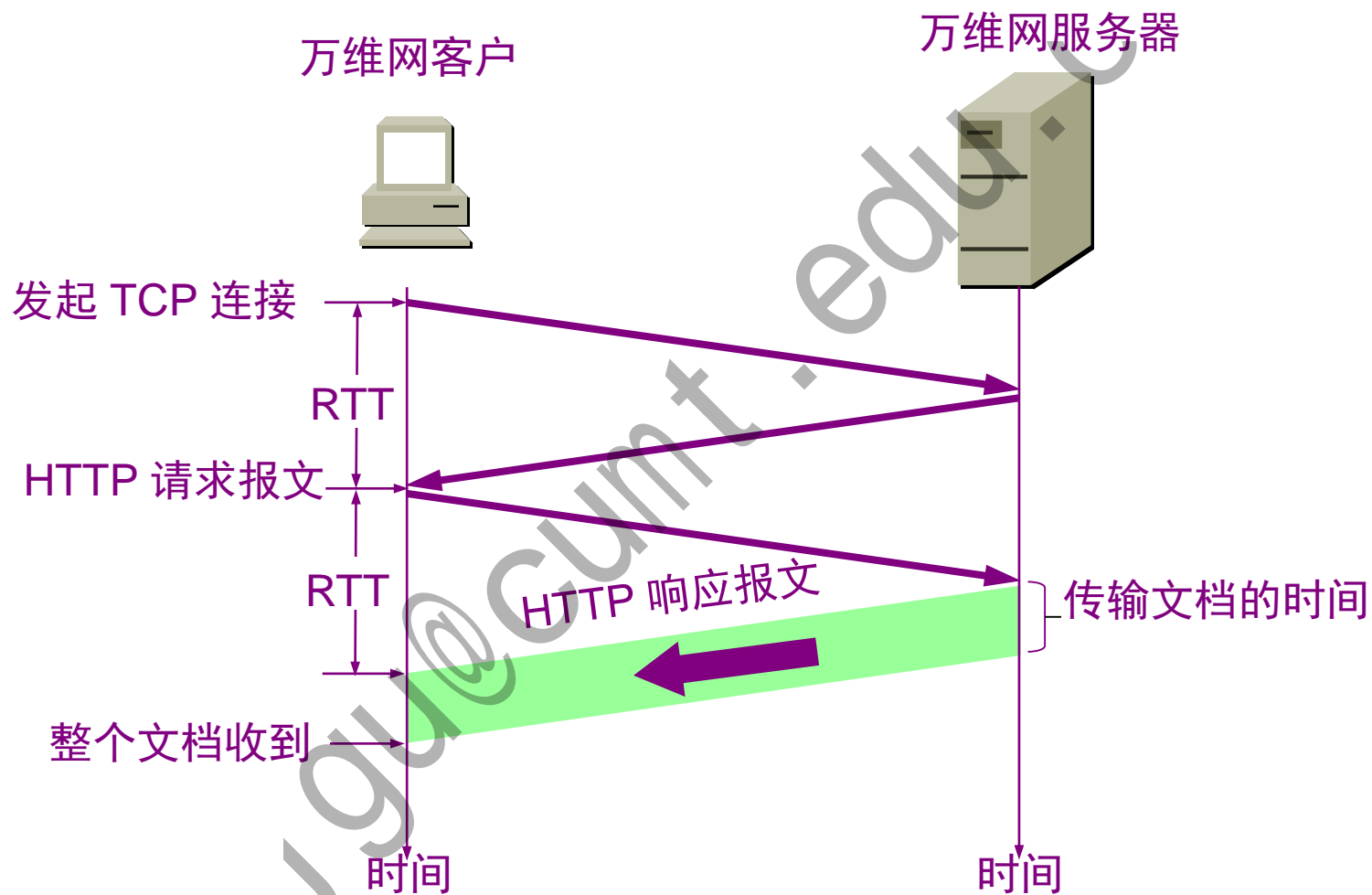
HTTP/1.0的主要缺点之一

- 每请求一个文档就要有两倍RTT的开销。
 - 若一个主页上的很多链接的对象(如图像等)需要依次进行链接, 那么每一次链接下载都导致 $2 \times \text{RTT}$ 的开销。



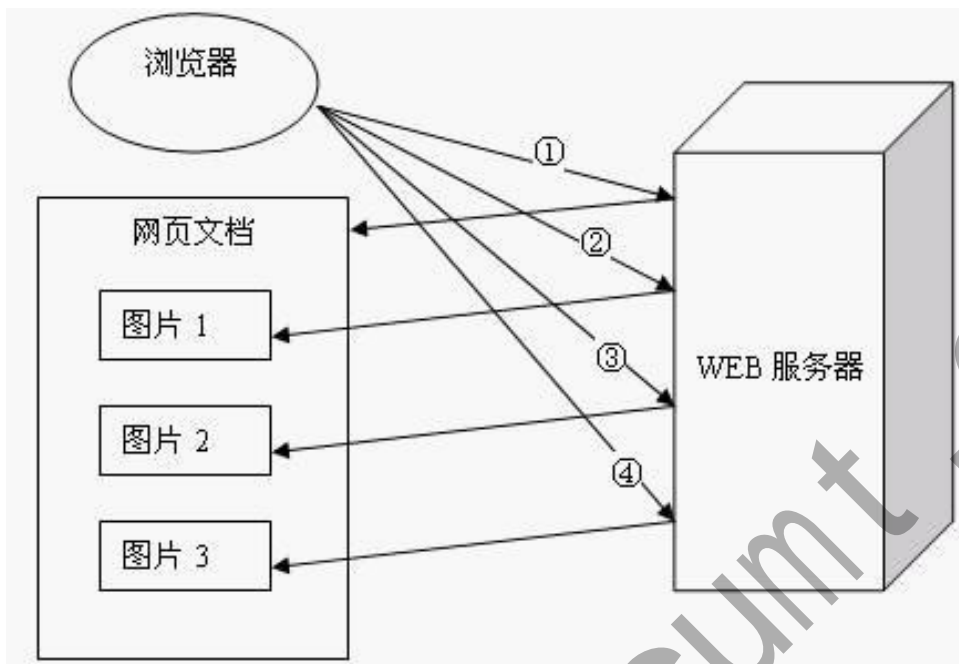


请求一个万维网文档所需的时间





一个万维网文档包含多个链接对象的情况



一个包含有许多图像的网页文件中并没有包含真正的图像数据内容，而只是指明了这些图像的URL地址。

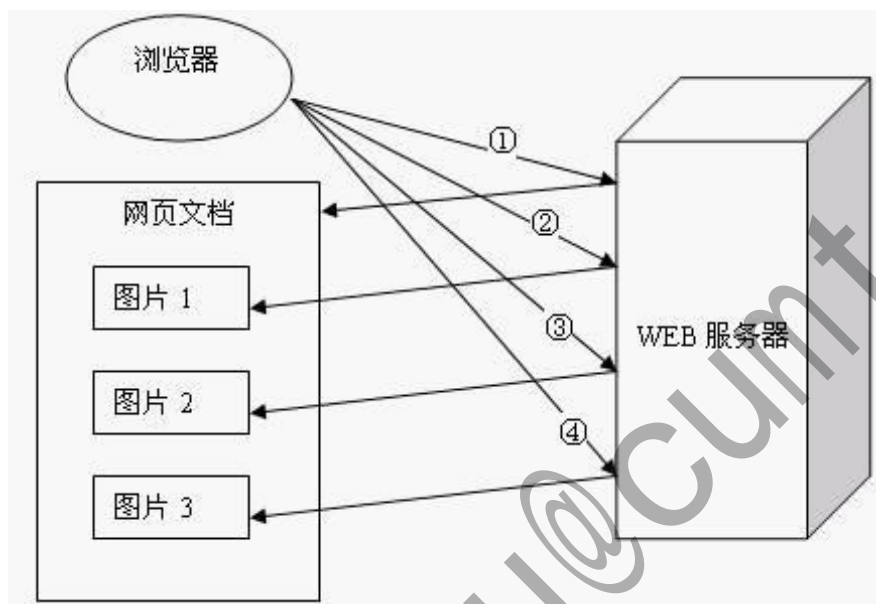
- ✓ 当WEB浏览器访问这个网页文件时，浏览器首先要发出针对该网页文件的请求；
- ✓ 当浏览器解析返回的该网页文档中的HTML内容时，发现其中的img图像标签后，将根据img标签中的src属性所指定的URL地址，再次向服务器发出下载图像数据的请求。





HTTP/1.0的主要缺点之二

- 万维网客户和服务端为每一次建立新的TCP连接都要分配缓存和变量。



访问一个包含有许多图像的网页文件的整个过程包含了多次请求和响应，每次请求和响应都需要建立一个单独的连接，每次连接只是传输一个文档或图像，上一次和下一次请求完全分离。

即使图像文件都很小，但是客户端和服务端每次建立和关闭连接却是一个相对比较费时的过程，并且会严重影响客户机和服务器的性能。





特别是万维网服务器往往要同时服务于大量客户的请求，所以这种非持续连接会使外围网服务器往往要同时服务于大量客户的请求，所以这种非持续连接会使万维网服务器的负担很重。



例题1:

- 假定在浏览器上点击一个URL，但是这个URL的IP地址以前并没有缓存在本地主机上。因此需要用DNS自动查找和解析。
- 假定要解析到所要找到的URL的IP地址共经过 n 个DNS服务器，所经过的时间分别为 $RTT_1, RTT_2, \dots, RTT_n$ 。
- 假定从要找的网页上只需读取一个很小的图片(即忽略这个小图片的传输时间)。从本地主机到这个网页的往返时间是 RTT_w 。
- 试问从点击这个URL开始，一直到本地主机的屏幕上出现所读取的小图片，一共要经过多少时间？

解答:

解析IP地址的时间: $RTT_1 + RTT_2 + \dots + RTT_n$

建立TCP连接和请求Web文档的时间: $2RTT_w$

需要的总时间:

$$2RTT_w + RTT_1 + RTT_2 + \dots + RTT_n$$





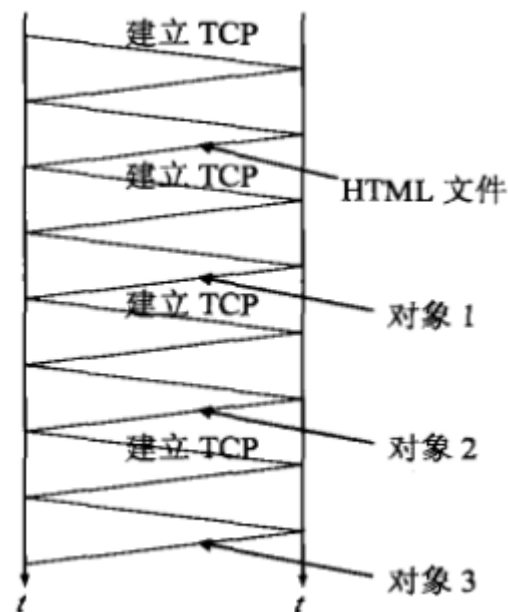
例题2:

- 在上题中，假定同一台服务器的HTML文件中又链接了三个非常小的对象。若忽略这些对象的发送时间，试计算客户点击读取这些对象所需的时间。
- (1) 没有并行TCP连接的非持续HTTP;
- (2) 使用并行TCP连接的非持续HTTP;

解答:

(1) 所需时间:

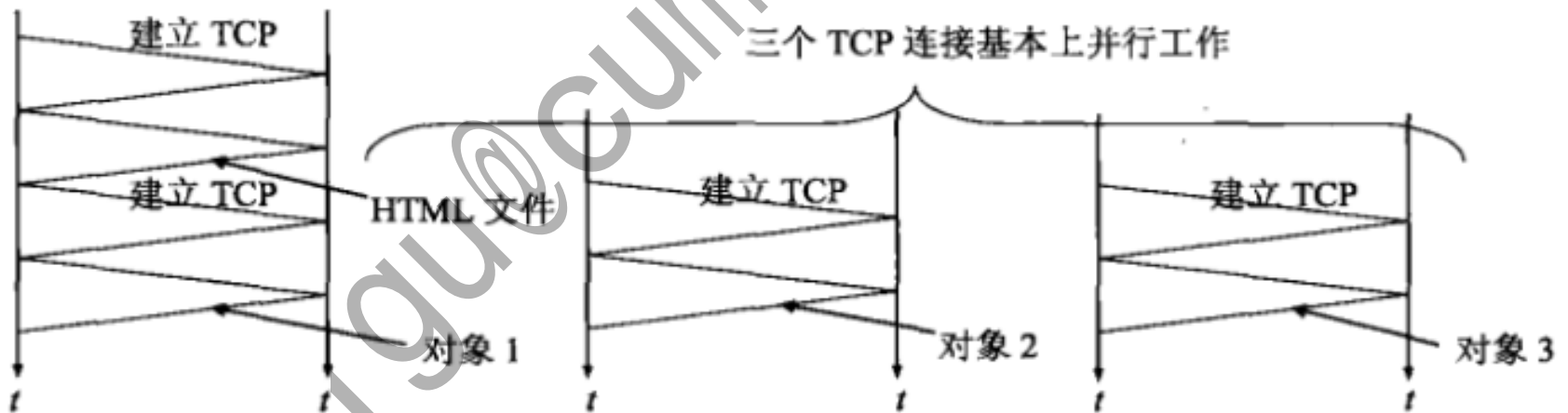
$$\begin{aligned} &= RTT_1 + RTT_2 + \dots + RTT_n \text{ (解析IP地址)} \\ &\quad + 2RTT_w \text{ (建立TCP连接和读取HTML文件)} \\ &\quad + 3 \times (2RTT_w) \text{ (读取三个对象)} \\ &= RTT_1 + RTT_2 + \dots + RTT_n + 8RTT_w \end{aligned}$$





(2) 所需时间:

$$\begin{aligned} &= \text{RTT}_1 + \text{RTT}_2 + \dots + \text{RTT}_n \text{ (解析IP地址)} \\ &\quad + 2\text{RTT}_w \text{ (建立TCP连接和读取HTML文件)} \\ &\quad + 2\text{RTT}_w \text{ (并行地建立TCP连接和并行地读取三个对象)} \\ &= \text{RTT}_1 + \text{RTT}_2 + \dots + \text{RTT}_n + 4\text{RTT}_w \end{aligned}$$





Q17: HTTP/1.1如何工作?

- HTTP /1.1支持**持续连接**(persistent connection), 在一个TCP连接上可以传送多个HTTP请求和响应, 减少了建立和关闭连接的消耗和延迟。一个包含有许多图像的网页文件的多个请求和应答可以在一个连接中传输, 但每个单独的网页文件的请求和应答仍然需要使用各自的连接。
- 这并不局限于传送同一个页面上链接的文档, 而是只要这些文档都在同一个服务器上就行。
- 目前一些流行的浏览器(例如, IE 6.0)的默认设置就是使用 HTTP/1.1。





持续连接的两种工作方式

- 非流水线方式：客户在收到前一个响应后才能发出下一个请求。这比非持续连接的两倍 RTT 的开销节省了建立 TCP 连接所需的一个 RTT 时间。但服务器在发送完一个对象后，其 TCP 连接就处于空闲状态，浪费了服务器资源。
- 流水线方式：客户在收到 $HTTP$ 的响应报文之前就能够接着发送新的请求报文。一个接一个的请求报文到达服务器后，服务器就可连续发回响应报文。使用流水线方式时，客户访问所有的对象只需花费一个 RTT 时间，使 TCP 连接中的空闲时间减少，提高了下载文档效率。





基于HTTP/1.1的信息交换过程

- HTTP/1.1允许客户端不用等待上一次请求结果返回，就可以发出下一次请求，但服务器端必须按照接收到客户端请求的先后顺序依次回送响应结果，以保证客户端能够区分出每次请求的响应内容，这样也显著地减少了整个下载过程所需要的时间。



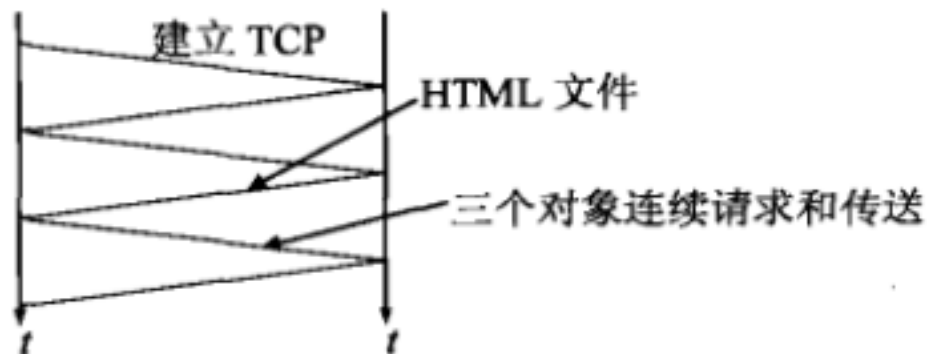


例题3:

- 在上题中，假定同一台服务器的HTML文件中又链接了三个非常小的对象。若忽略这些对象的发送时间，试计算客户点击读取这些对象所需的时间。
- (1) 没有并行TCP连接的非持续HTTP;
- (2) 使用并行TCP连接的非持续HTTP;
- (3) 流水线方式的持续HTTP。

(3) 所需时间:

$$\begin{aligned} &= \text{RTT}_1 + \text{RTT}_2 + \dots + \text{RTT}_n \text{ (解析IP地址)} \\ &\quad + 2\text{RTT}_w \text{ (建立TCP连接和读取HTML文件)} \\ &\quad + \text{RTT}_w \text{ (连续读取三个对象)} \\ &= \text{RTT}_1 + \text{RTT}_2 + \dots + \text{RTT}_n \\ &\quad + 3\text{RTT}_w \end{aligned}$$





例题3:

如果改为“链接对象”，那么每个页面中包含的对象个数会影响检索时间。

- 一个万维网网点有1千万个页面，平均每个页面有10个超链。读取一个页面平均要100ms。问要检索整个网点所需的最少时间？

解答：因为题目未指出是否还要点击这10个超链，也没有给出点击一个超链需要多少时间，以及是否要在点击超链后再继续点击下去……，所以一个页面上有10个超链，和本题的解答没有关系。

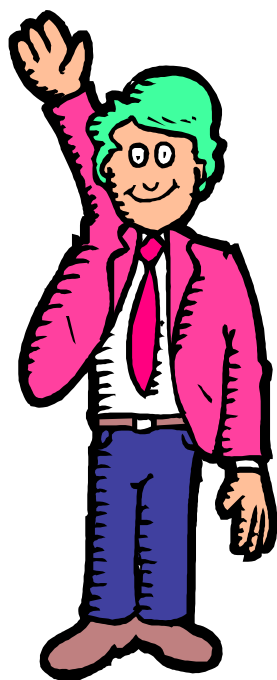
读取一个页面平均要 100 ms，那么读取 1000 万个页面需要的时间：

$$T = 1000 \times 10^4 \times 100 \times 10^{-3} = 10^6 \text{ s}$$

$$1 \text{ 天} = 24 \times 3600 \text{ s} = 86400 \text{ s}$$

$$T \approx 11.57 \text{ 天}$$





**THANK
YOU!**

