

第七章 应用层

万维网



万维网 (World Wide Web)

- 1989年，Web (也叫 WWW)诞生于欧洲原子能中心 CERN
- 1994年，CERN 和 M.I.T. 签署了建立万维网联盟 (World Wide Web Consortium, W3C) 的协议，致力于进一步开发 Web、协议标准化和站点互操作等
- Web 是web网页的集合 (collection of web pages)
- 每个页面包含了指向其他页面的链接 (超级链接)
- 浏览器 –显示阅读web页面的程序



Tim Berners-Lee

- Web之父
- 出生数学世家，英国计算机科学家
- 1976，牛津大学物理学学士；后在公司担任 coder；1980年，到瑞士日内瓦CERN工作
- 1989年仲夏之夜，蒂姆成功开发出世界上第一个Web服务器和第一个Web客户机



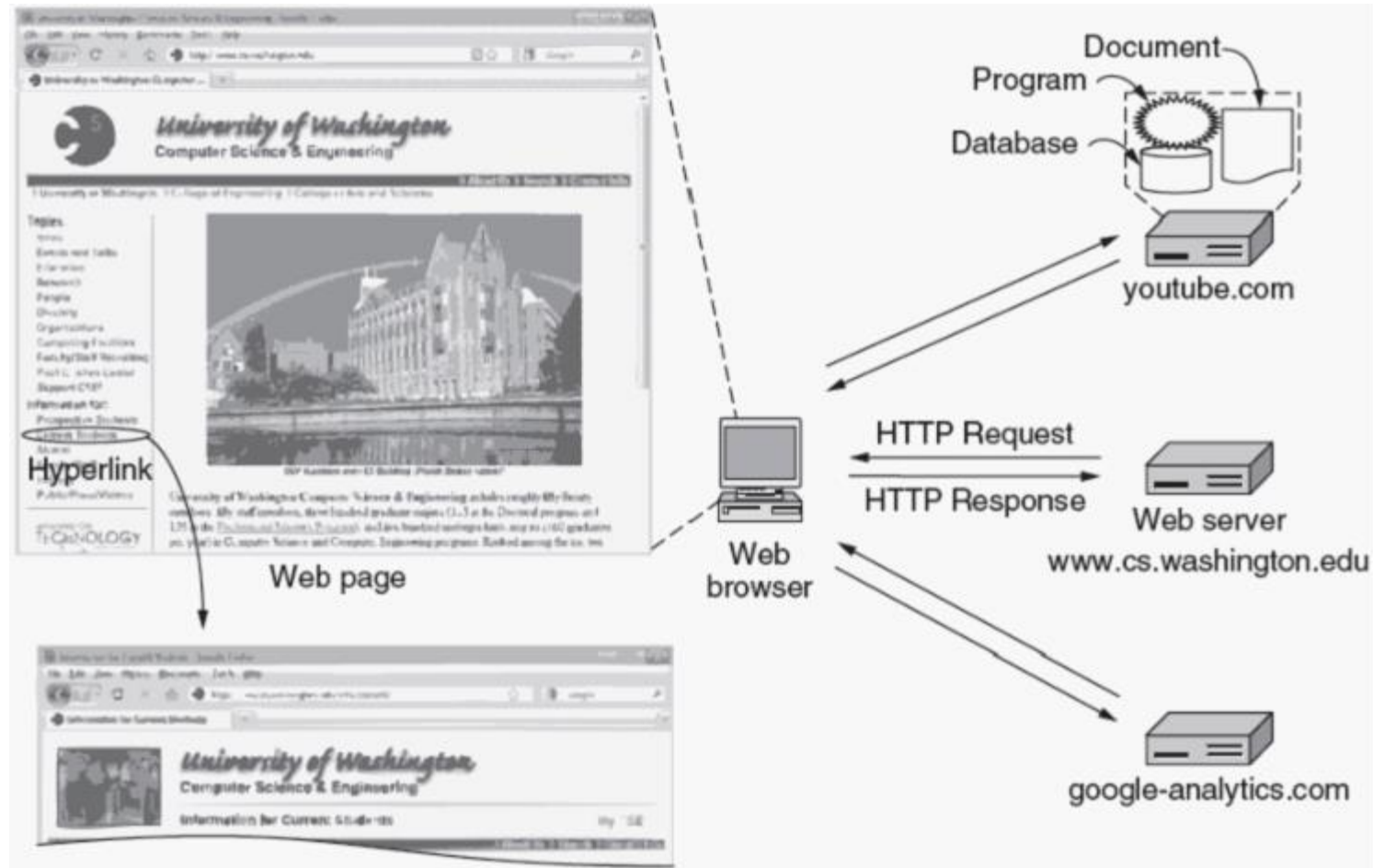


WWW的组成部分

- 资源，web页面，Resource (html)
- 统一资源定位器：URL
- 通信协议HTTP
- 比较WWW和互联网（internet）



Web的体系结构





Web的体系结构

□ 怎样找到页面？

➤ URLs (Uniform Resource Locators)

□ 例： `http://www.phdcomics.com/comics.php`

Protocol

Server

Page on server

Name	Used for	Example
http	Hypertext (HTML)	http://www.ee.uwa.edu/~rob/
https	Hypertext with security	https://www.bank.com/accounts/
ftp	FTP	ftp://ftp.cs.vu.nl/pub/minix/README
file	Local file	file:///usr/suzanne/prog.c
mailto	Sending email	mailto:JohnUser@acm.org
rtsp	Streaming media	rtsp://youtube.com/montypython.mpg
sip	Multimedia calls	sip:eve@adversary.com
about	Browser information	about:plugins



□ Web页面由 **URL (Uniform Resource Locators)标识** (i.e.

`http://www.abcd.com/products.html`)

➤ 协议: `http`

➤ 页面所在的机器的DNS 域名: `www.abcd.com`

➤ 包含web页面的文件的名称: `products.html`



- 当用户单击一个超级链接（URL）时：
 - 浏览器检查URL (读取浏览器的输入)
 - 浏览器向 DNS 服务器询问域名的IP地址
 - DNS 返回对应的 IP 地址
 - 浏览器和Web服务器建立TCP 连接（在端口 80）
 - 浏览器发送请求，要求获取文件products.html
 - Web服务器返回被请求的文件
 - TCP 连接被释放
 - 浏览器解释显示下载到本地的文件



- 一个web页面可能由PDF文件、GIF图标、MPEG视频、MP3歌曲，或者其他数百种文件类型的任何一种组成
- 浏览器可能在解释这些文件的时候会遇到问题，不是让浏览器越来越大，而是采用了一个更加通用的解决方案
- 当服务器返回一个页面，它通常也返回一些有关该页面的附加信息，包含了页面的MIME类型，以决定如何显示该页面



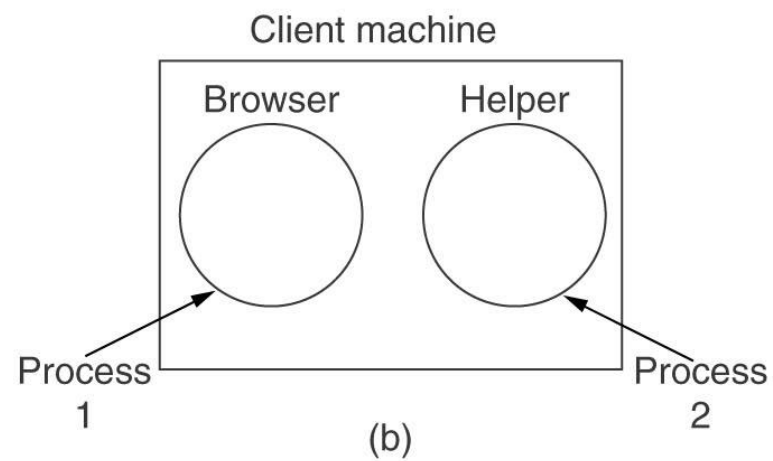
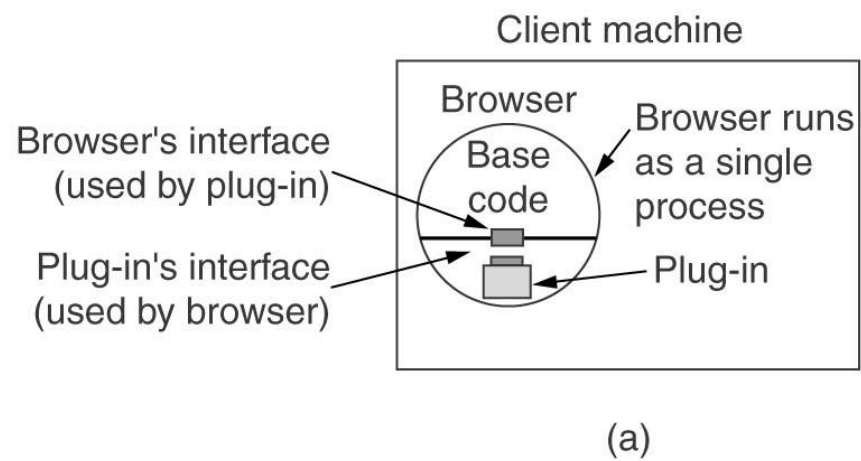
□ 有两种可能的扩展浏览器的方式

Plug-ins

代码模块，运行在浏览器的内部

Helper
applications

独立的程序，浏览器只是把参数传入





服务器端

- 典型的web 服务器的操作：
 - 接收来自客户的TCP连接
 - 获取所需文件的名字
 - 从本地磁盘上获取文件（静态页面）
 - 将文件返回给客户
 - 释放TCP连接



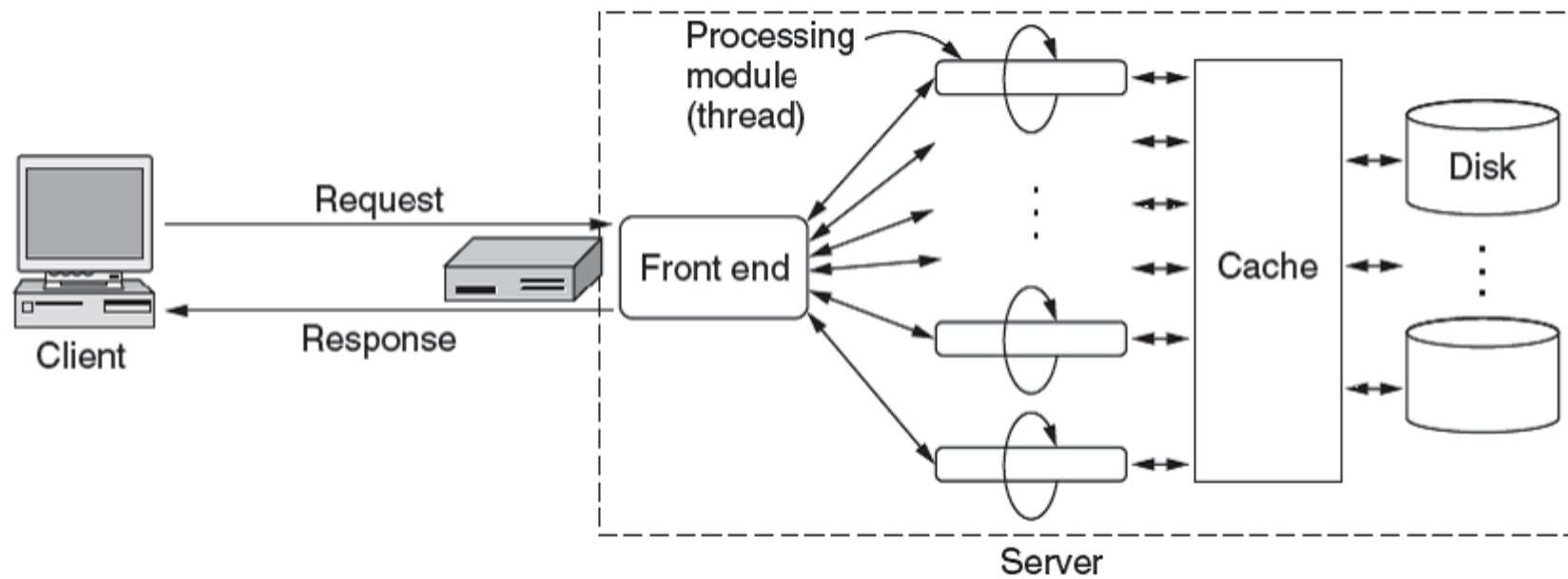
服务器端

□ 改进

- 在内存维护一个缓存，保存最近用过的 n 个文件
- 多线程服务器



多线程web服务器

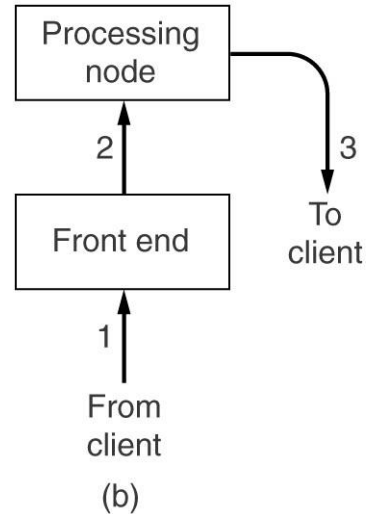
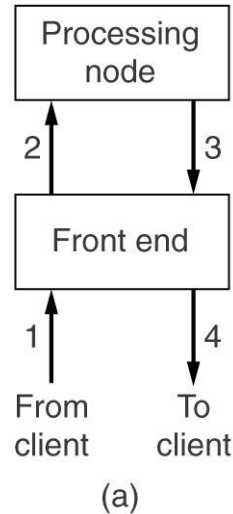




服务器端

□ 客户的TCP连接中止于前端，所以应答也必须经过前端 (a)

➤ 一种解决的方法是**TCP移交**，TCP端点被传递给处理节点，所以应答可以直接向客户端发送 (b)



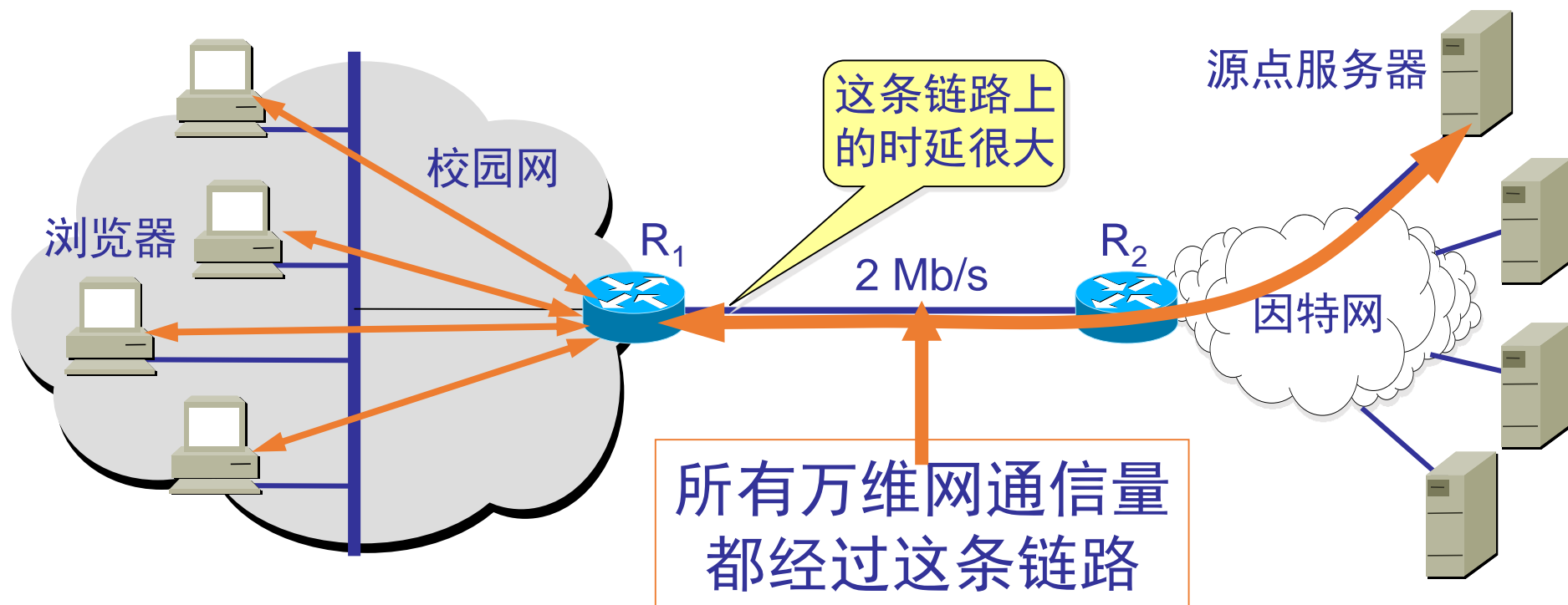


高速代理服务器

- 代理服务器(proxy server)又称为万维网高速缓存(Web cache), 它代表浏览器发出 HTTP 请求。
- 万维网高速缓存把最近的一些请求和响应暂存在本地磁盘中。
- 当与暂时存放的请求相同的新请求到达时, 万维网高速缓存就把暂存的响应发送出去, 而不需要按 URL 的地址再去因特网访问该资源。

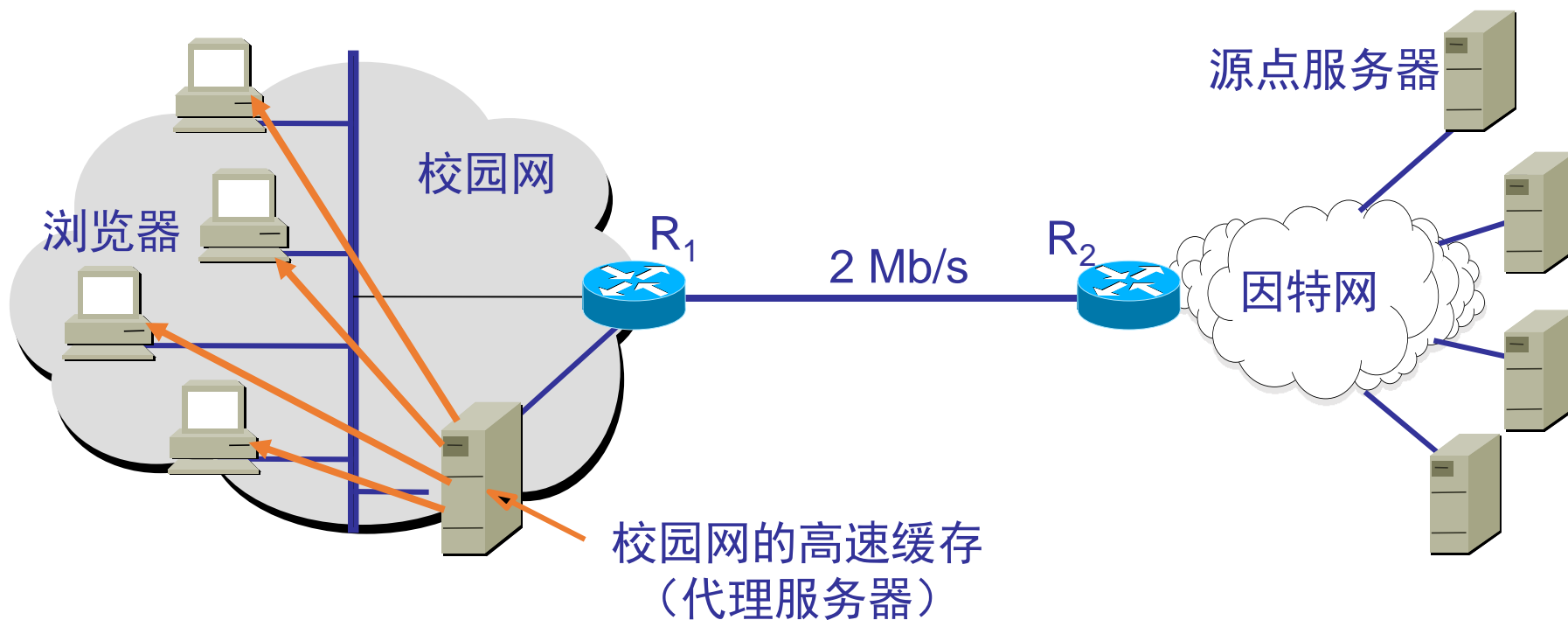
如果没有高速缓存……

No proxy server……



有了高速缓存……

Have a proxy server……





问题

- 不同的用户登录同一个web服务器，想获得不同的资源：个性化服务
- 一个电子商城，服务器如何跟踪购物车上的项目？
- 一个门户网站，如何向定制的用户推送不同的信息？
- 可能的答案：IP地址

不可行



- 一个小于4kB的命名串
- 当客户请求时，web服务器除了应答外，附送一个cookie，存储在客户机磁盘
- 客户再访问同一个web服务器时，同时发送cookie
- 服务器辨识出该用户，并得到它关心的一些信息
- 方便的同时，是否侵犯了用户的隐私？



小结

- WWW的三个组成部分
 - 资源
 - 资源定位器
 - 传输协议
- WWW的传输层使用了TCP
- 浏览器解释资源的不同方式：内嵌插件或外挂
- TCP移交缓解了前端的压力

思考题

- 什么是万维网？
- 万维网的主要组成部分有哪些？
- URL是怎样构成的？
- 为什么需要TCP移交？它解决了什么问题？
- 热门网站如何应对众多客户请求？
- 什么是COOKIE？

谢谢观看

致谢

本课程课件中的部分素材来自于：（1）清华大学出版社出版的翻译教材《计算机网络》（原著作者：Andrew S. Tanenbaum, David J. Wetherall）；（2）思科网络技术学院教程；（3）网络上搜到的其他资料。在此，对清华大学出版社、思科网络技术学院、人民邮电出版社、以及其它提供本课程引用资料的个人表示衷心的感谢！

对于本课程引用的素材，仅用于课程学习，如有任何问题，请与我们联系！