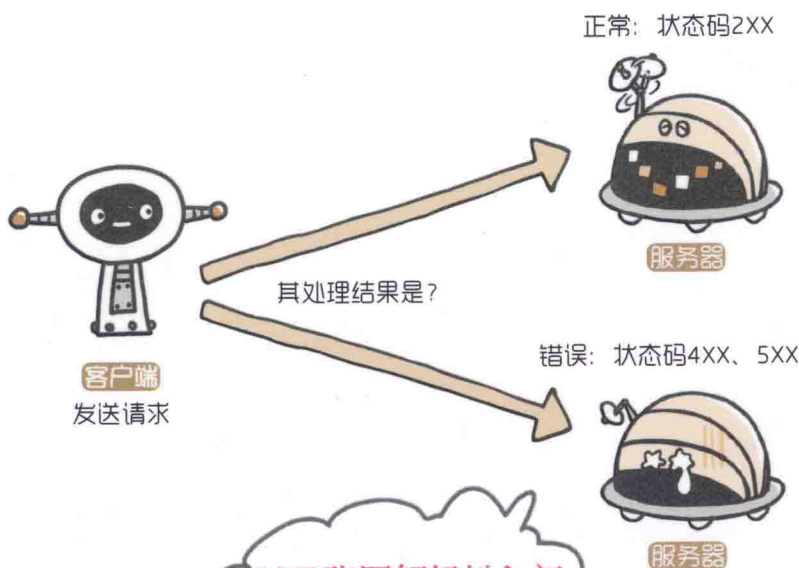


TURING

图灵程序
设计丛书

图解HTTP

OWASP 日本分会主席 【日】上野宣 著
于均良 译



172张图解轻松入门

从基础知识到最新动向
一本书掌握HTTP协议

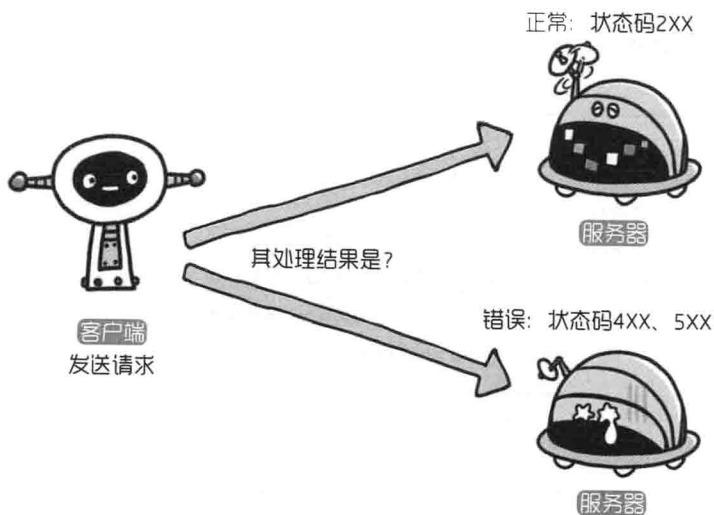
人民邮电出版社
POSTS & TELECOM PRESS

TURING

图灵程序
设计丛书

图解HTTP

OWASP
日本分会主席 【日】上野宣 著
于均良 译



人民邮电出版社
北京

图书在版编目(CIP)数据

图解HTTP/(日)上野宣著;于均良译.--北京:
人民邮电出版社,2014.5
(图灵程序设计丛书)
ISBN 978-7-115-35153-1

I. ①图… II. ①上… ②于… III. ①计算机网络—
通信协议 IV. ①TN915.04

中国版本图书馆CIP数据核字(2014)第055492号

内 容 提 要

本书对互联网基石——HTTP协议进行了全面系统的介绍。作者由HTTP协议的发展历史娓娓道来,严谨细致地剖析了HTTP协议的结构,列举诸多常见通信场景及实战案例,最后延伸到Web安全、最新技术动向等方面。本书的特色为在讲解的同时,辅以大量生动形象的通信图例,更好地帮助读者深刻理解HTTP通信过程中客户端与服务端之间的交互情况。读者可通过本书快速了解并掌握HTTP协议的基础,前端工程师分析抓包数据,后端工程师实现REST API、实现自己的HTTP服务器等过程中所需的HTTP相关知识本书均有介绍。

本书适合Web开发工程师,以及对HTTP协议感兴趣的各层次读者。

◆ 著 [日] 上野 宣

译 于均良

责任编辑 徐 骞

责任印制 焦志炜

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京天宇星印刷厂印刷

◆ 开本: 880×1230 1/32

印张: 8.25

字数: 220千字 2014年5月第1版

印数: 1-4000册 2014年5月北京第1次印刷

著作权合同登记号 图字: 01-2013-8984号

定价: 49.00元

读者服务热线: (010)51095186 转 600 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京崇工商广字第0021号



版 权 声 明

HTTP の教科書 (ISBN 978-4-7981-2625-8)

Copyright © 2013 by SEN UENO.

Original Japanese edition published by SHOEISHA Co., Ltd

Simplified Chinese Character translation rights arranged with
SHOEISHA Co., Ltd

through CREEK & RIVER Co., Ltd

Simplified Chinese Character translation copyright © 2014 by Posts &
Telecom Press

本书中文简体字版由 SHOEISHA Co., Ltd. 授权人民邮电出版社独家出版。未经出版者书面许可，不得以任何方式复制或抄袭本书内容。版权所有，侵权必究。

(图灵公司感谢李典对本书的审读。)

译者序

目前，国内讲解 HTTP 协议的书实在太少了。

在我的印象中，讲解网络协议的书仅有两本。一本是《HTTP 权威指南》，但其厚度令人望而生畏；另一本是《TCP/IP 详解，卷 1》，内容艰涩难懂，学习难度较大。这两本书都是被读者们奉为“圣经”的经典之作，大师们的授道自然无可挑剔，但关键是它们对初学者都不那么友好，大家的学习信心很容易受到打击，阅读中途或将束之高阁。本书的出现及时缓解了该问题。

HTTP 协议本身并不复杂，理解起来也不会花费太多学习成本，但纯概念式的学习稍显单调。前端工程师也许对各种具有炫酷效果的页面的实现技巧、赏心悦目的 UI 框架更感兴趣，但因此常常忽视了 HTTP 协议这部分基础内容。实际上，如果想要在专业技术道路上走得更坚实，绝对不能绕开学习 HTTP 协议这一环节。对基础及核心部分的深入学习是成为一名专业技术人员的前提，以不变应万变才是立足之本。

我在学习 Web 开发的过程中，曾接触到编写网络爬虫程序、分析抓包数据、实现 HTTP 服务器、提供网站 REST API、修改后端定制框架等方面，它们无一例外，都会用到 HTTP 协议的各方面知识，并且某些细节无法通过查阅资料立即领会到，还需依靠扎实的基础及平日里的积累。

本书作者的写作手法平实易懂，内容讲解透彻到位。前半部分由 HTTP 的成长发展史娓娓道来，基于 HTTP 1.1 标准讲解通信过程，包括 HTTP 方法、协议格式、报文结构、首部字段、状态码等的具体含义，还分别讲解 HTTP 通信过程中代理、网关、隧道等的作用。接着介绍 SPDY、WebSocket、WebDAV 等 HTTP 的扩展功能。作者还从细节方面举例，让读者更好地理解何为无状态（stateless）、301 和 302 重定向的区别在哪、缓存机制，等等。本书后半部分的重心放在 Web 安全上，涵盖 HTTPS、SSL、证书认证、加密机制、Web 攻击手段等内容。

旨在让读者对 HTTP 协议形成一个整体概念，明确设计 HTTP 的目的及意义所在，了解 HTTP 的工作机制，掌握报文中常用的首部字段，返回结果状态码的作用，对各种客户端与服务器的通信交互场景的细节等都做到了了然于心，从而在平时的开发工作中独立思考，迅速准确地定位分析由 HTTP 引发的问题，并辅以适当的方法加以解决。

本书图文并茂，大量图片穿插文中，生动形象地向读者介绍每一个应用案例，减少了读者阅读时的枯燥感。借助一张张配图，读者们不仅会加深视觉记忆，在轻松愉悦的氛围中，还可以更深刻地理解通信机制等背后的工作原理。正所谓一图胜千文。

在本书即将付梓之际，感谢 EMC 首席工程师高博学长、IBM 工程师李亚舟 (Fleuria)、豆瓣运维工程师钱龙、全栈工程师缪思源 (Aveline Swan) 以及姜莹等好友。他们在繁重的工作之余，牺牲个人闲暇时间，耐心地帮我扫清技术疑点，修正翻译疏漏，在此谨表示衷心的感谢。

最后祝大家阅读愉快！

于均良

2014年1月

前言

本书的上一版是 2004 年出版的《今夜わかる HTTP》(中文译名:今晚我们一起学习 HTTP,翔泳社)。和当时一样,现在互联网的主流仍是 Web,但人们对 Web 的要求却不断地发生变化。Google 在 2005 年推出了地图服务 Google Maps,很多人看到这一 Web 应用程序的界面后感到十分震惊。因为在此之前,我们只能借助桌面应用程序或 Flash 等方式,实现流畅滚动及视角放大缩小等功能,如今这些功能仅需一个 Web 浏览器就能呈现了。也许正是由于 Google Maps 的出现,人们对 Web 的要求才开始变得多了起来。发送请求、等待响应,这些 HTTP 中稀松平常的功能已经无法满足人们的需求了。于是,Web 不再停留在 HTTP/1.1 版本,在保持 HTTP 简洁的同时,也开始开发新的功能。我之所以要撰写《今夜わかる HTTP》一书,是因为我发现多数 Web 应用程序开发者并不了解支撑 Web 基础的 HTTP 协议。我坚信通过学习协议,大家能更深刻地理解 Web 开发。即使是在本书撰写完成后的今天,我的这一想法仍未改变,肯定还有很多开发者尚未了解 HTTP 协议。

对 HTTP 协议有了更深入的理解后,也许你会从中得到一些启发。不再囿于 HTTP/1.1 版本的制约,你也能开发出 Google Maps 那样的应用程序。

本书不仅面向 Web 应用程序的开发者,还面向使用 Web 的软件开发、Web 风险评估的安全工程师、前端工程师以及 Web 使用者等与 Web 相关的所有读者,希望这本书能对大家有所帮助。

写于华盛顿 DC 的酒店

2013 年 1 月吉日

TRICORDER 株式会社 上野宣

致谢

Masato Kinugawa 先生

感谢您细致的检查核对，并站在读者的角度提出宝贵建议。当您指出我书中的 MD5 hash 值不是 abcd，而是 abc 时，我感到十分惊讶，感叹您不愧是世界级权威专家。能和您一起工作不胜荣幸。

山崎圭吾先生

感谢您及时帮我检查并提出真知灼见。每当我对稿件有所改动，您总是立即着手帮我核对。我曾担心这样是否会对您的本职工作造成影响，不过我就不在此和您客气了，总之十分感谢您的帮助。

viii

Netagent 株式会社 长谷川阳介先生

您曾对我说过：“不感兴趣的领域我不了解，但我擅长的领域就放心交给我吧！”感谢您以自己独特的方式对本书进行核查。作为安全领域里首屈一指的权威专家，您能协助本书进行审读工作，我倍感荣幸。

我身边的朋友们

在我撰写本书时，我常常顾不上工作，有时又疏于写作，给大家添了很多麻烦。正是得到了诸位的理解与支持，本书才得以顺利出版。在此，我要特别感谢本书的编辑野村先生。

目录

第1章 了解Web及网络基础 001

- 1.1 使用HTTP协议访问Web002
- 1.2 HTTP的诞生003
 - 1.2.1 为知识共享而规划Web 003
 - 1.2.2 Web成长时代 004
 - 1.2.3 驻足不前的HTTP 005
- 1.3 网络基础TCP/IP006
 - 1.3.1 TCP/IP协议族 006
 - 1.3.2 TCP/IP的分层管理 007
 - 1.3.3 TCP/IP通信传输流 009
- 1.4 与HTTP关系密切的协议: IP、TCP和DNS010
 - 1.4.1 负责传输的IP协议 011
 - 1.4.2 确保可靠性的TCP协议 012
- 1.5 负责域名解析的DNS服务013
- 1.6 各种协议与HTTP协议的关系014
- 1.7 URI和URL.....016
 - 1.7.1 统一资源标识符 016
 - 1.7.2 URI格式 017

第2章 简单的HTTP协议 021

- 2.1 HTTP协议用于客户端和服务器端之间的通信022
- 2.2 通过请求和响应的交换达成通信.....022
- 2.3 HTTP是不保存状态的协议025
- 2.4 请求URI定位资源.....026
- 2.5 告知服务器意图的HTTP方法027
- 2.6 使用方法下达命令.....033
- 2.7 持久连接节省通信量.....034
 - 2.7.1 持久连接 036
 - 2.7.2 管线化 037
- 2.8 使用Cookie的状态管理037

第3章 HTTP报文内的HTTP信息 041

- 3.1 HTTP报文042



3.2	请求报文及响应报文的结构	042
3.3	编码提升传输速率	044
3.3.1	报文主体和实体主体的差异	044
3.3.2	压缩传输的内容编码	044
3.3.3	分割发送的分块传输编码	045
3.4	发送多种数据的多部分对象集合	046
3.5	获取部分内容的范围请求	048
3.6	内容协商返回最合适的内容	050

第4章 返回结果的HTTP状态码 053

4.1	状态码告知从服务器端返回的请求结果	054
4.2	2XX 成功	055
4.2.1	200 OK	055
4.2.2	204 No Content	056
4.2.3	206 Partial Content	056
4.3	3XX 重定向	056
4.3.1	301 Moved Permanently	057
4.3.2	302 Found	057
4.3.3	303 See Other	058
4.3.4	304 Not Modified	059
4.3.5	307 Temporary Redirect	059
4.4	4XX 客户端错误	060
4.4.1	400 Bad Request	060
4.4.2	401 Unauthorized	060
4.4.3	403 Forbidden	061
4.4.4	404 Not Found	061
4.5	5XX 服务器错误	062
4.5.1	500 Internal Server Error	062
4.5.2	503 Service Unavailable	062

第5章 与HTTP协作的Web服务器 065

5.1	用单台虚拟主机实现多个域名	066
5.2	通信数据转发程序：代理、网关、隧道	067
5.2.1	代理	068
5.2.2	网关	070
5.2.3	隧道	070
5.3	保存资源的缓存	071
5.3.1	缓存的有效期限	072

5.3.2 客户端的缓存	072
--------------------	-----

第6章 HTTP 首部 075

6.1 HTTP 报文首部	076
6.2 HTTP 首部字段	078
6.2.1 HTTP 首部字段传递重要信息	078
6.2.2 HTTP 首部字段结构	078
6.2.3 4种HTTP首部字段类型	079
6.2.4 HTTP/1.1 首部字段一览	080
6.2.5 非HTTP/1.1首部字段	082
6.2.6 End-to-end 首部和Hop-by-hop 首部	083
6.3 HTTP/1.1 通用首部字段	083
6.3.1 Cache-Control	084
6.3.2 Connection	091
6.3.3 Date	093
6.3.4 Pragma	094
6.3.5 Trailer	095
6.3.6 Transfer-Encoding	096
6.3.7 Upgrade	097
6.3.8 Via	098
6.3.9 Warning	099
6.4 请求首部字段	100
6.4.1 Accept	101
6.4.2 Accept-Charset	102
6.4.3 Accept-Encoding	103
6.4.4 Accept-Language	104
6.4.5 Authorization	105
6.4.6 Expect	106
6.4.7 From	107
6.4.8 Host	107
6.4.9 If-Match	108
6.4.10 If-Modified-Since	110
6.4.11 If-None-Match	111
6.4.12 If-Range	112
6.4.13 If-Unmodified-Since	113
6.4.14 Max-Forwards	114
6.4.15 Proxy-Authorization	115
6.4.16 Range	116
6.4.17 Referer	116
6.4.18 TE	117



6.4.19	User-Agent	118
6.5	响应首部字段	119
6.5.1	Accept-Ranges	119
6.5.2	Age	120
6.5.3	ETag	120
6.5.4	Location	122
6.5.5	Proxy-Authenticate	123
6.5.6	Retry-After	123
6.5.7	Server	124
6.5.8	Vary	125
6.5.9	WWW-Authenticate	125
6.6	实体首部字段	126
6.6.1	Allow	126
6.6.2	Content-Encoding	127
6.6.3	Content-Language	128
6.6.4	Content-Length	128
6.6.5	Content-Location	129
6.6.6	Content-MD5	129
6.6.7	Content-Range	130
6.6.8	Content-Type	131
6.6.9	Expires	131
6.6.10	Last-Modified	132
6.7	为Cookie服务的首部字段	132
6.7.1	Set-Cookie	134
6.7.2	Cookie	136
6.8	其他首部字段	137
6.8.1	X-Frame-Options	137
6.8.2	X-XSS-Protection	138
6.8.3	DNT	138
6.8.4	P3P	139

第7章 确保Web安全的HTTPS **141**

7.1	HTTP的缺点	142
7.1.1	通信使用明文可能会被窃听	142
7.1.2	不验证通信方的身份就可能遭遇伪装	146
7.1.3	无法证明报文完整性,可能已遭篡改	148
7.2	HTTP+加密+认证+完整性保护=HTTPS	150
7.2.1	HTTP加上加密处理和认证以及完整性保护后即是HTTPS	150
7.2.2	HTTPS是身披SSL外壳的HTTP	151

7.2.3	相互交换密钥的公开密钥加密技术	152
7.2.4	证明公开密钥正确性的证书	155
7.2.5	HTTPS的安全通信机制	161

第8章 确认访问用户身份的认证 167

8.1	何为认证.....	168
8.2	BASIC认证	169
8.3	DIGEST认证	171
8.4	SSL客户端认证.....	173
8.4.1	SSL客户端认证的认证步骤.....	174
8.4.2	SSL客户端认证采用双因素认证.....	175
8.4.3	SSL客户端认证必要的费用	175
8.5	基于表单认证.....	175
8.5.1	认证多半为基于表单认证	176
8.5.2	Session管理及Cookie应用.....	177

第9章 基于HTTP的功能追加协议 179

9.1	基于HTTP的协议	180
9.2	消除HTTP瓶颈的SPDY	180
9.2.1	HTTP的瓶颈	180
9.2.2	SPDY的设计与功能	184
9.2.3	SPDY消除Web瓶颈了吗	185
9.3	使用浏览器进行全双工通信的WebSocket	186
9.3.1	WebSocket的设计与功能	186
9.3.2	WebSocket协议	186
9.4	期盼已久的HTTP/2.0	189
9.5	Web服务器管理文件的WebDAV	190
9.5.1	扩展HTTP/1.1的WebDAV	191
9.5.2	WebDAV内新增的方法及状态码	192

第10章 构建Web内容的技术 195

10.1	HTML	196
10.1.1	Web页面几乎全由HTML构建	196
10.1.2	HTML的版本	197
10.1.3	设计应用CSS	198
10.2	动态HTML	198
10.2.1	让Web页面动起来的动态HTML	198



10.2.2	更易控制HTML的DOM	198
10.3	Web应用	200
10.3.1	通过Web提供功能的Web应用	200
10.3.2	与Web服务器及程序协作的CGI	200
10.3.3	因Java而普及的Servlet	201
10.4	数据发布的格式及语言	203
10.4.1	可扩展标记语言	203
10.4.2	发布更新信息的RSS/Atom	204
10.4.3	JavaScript衍生的轻量级易用JSON	206

第11章 Web的攻击技术

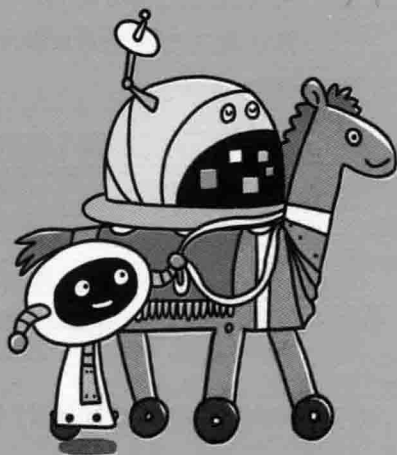
207

11.1	针对Web的攻击技术	208
11.1.1	HTTP不具备必要的安全功能	208
11.1.2	在客户端即可篡改请求	209
11.1.3	针对Web应用的攻击模式	210
11.2	因输出值转义不完全引发的安全漏洞	212
11.2.1	跨站脚本攻击	213
11.2.2	SQL注入攻击	218
11.2.3	OS命令注入攻击	223
11.2.4	HTTP首部注入攻击	225
11.2.5	邮件首部注入攻击	228
11.2.6	目录遍历攻击	229
11.2.7	远程文件包含漏洞	230
11.3	因设置或设计上的缺陷引发的安全漏洞	232
11.3.1	强制浏览	232
11.3.2	不正确的错误消息处理	234
11.3.3	开放重定向	237
11.4	因会话管理疏忽引发的安全漏洞	237
11.4.1	会话劫持	238
11.4.2	会话固定攻击	239
11.4.3	跨站点请求伪造	241
11.5	其他安全漏洞	242
11.5.1	密码破解	242
11.5.2	点击劫持	247
11.5.3	DoS攻击	249
11.5.4	后门程序	250

Chapter 1

第 1 章 了解 Web 及网络基础

本章概述了 Web 是建立在何种技术之上，以及 HTTP 协议是如何诞生并发展的。我们从其背景着手，来深入了解这部分内容。





1.1 使用 HTTP 协议访问 Web

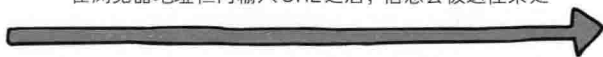
你知道当我们在网页浏览器（Web browser）的地址栏中输入 URL 时，Web 页面是如何呈现的吗？

当在浏览器的地址栏内输入 URL 时，可以看到 Web 页面
当然，即使你不了解其运作原理，也能看到 Web 页面

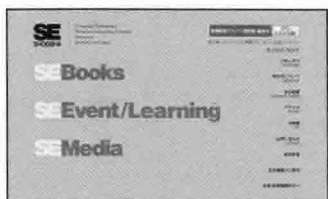


客户端

在浏览器地址栏内输入 URL 之后，信息会被送往某处



然后从某处获得的回复，内容就会显示在 Web 页面上



002

Web 页面当然不能凭空显示出来。根据 Web 浏览器地址栏中指定的 URL，Web 浏览器从 Web 服务器端获取文件资源（resource）等信息，从而显示出 Web 页面。

像这种通过发送请求获取服务器资源的 Web 浏览器等，都可称为客户端（client）。

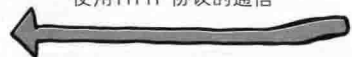


客户端

通过指定的访问地址获取(或上传)服务器资源(文件等信息)



使用 HTTP 协议的通信



服务器

Web 使用一种名为 HTTP（HyperText Transfer Protocol，超文本传输

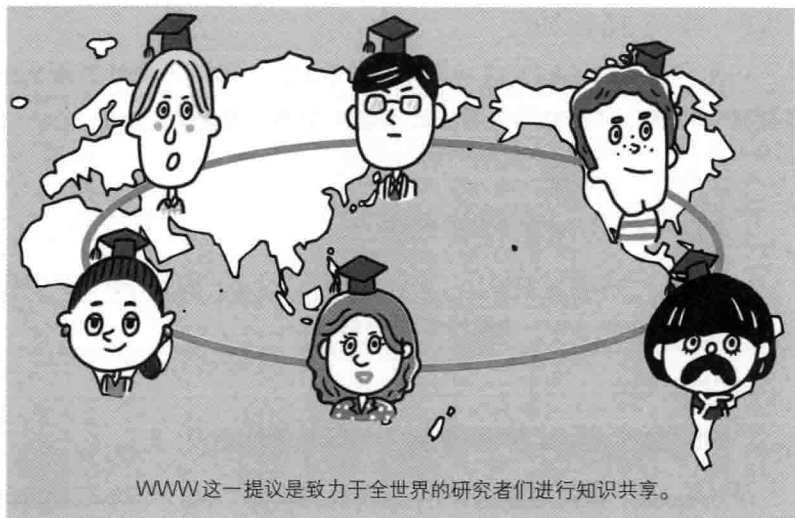
协议^①)的协议作为规范,完成从客户端到服务器端等一系列运作流程。而协议是指规则的约定。可以说,Web是建立在HTTP协议上通信的。

1.2 HTTP 的诞生

在深入学习HTTP之前,我们先来介绍一下HTTP诞生的背景。了解背景的同时也能了解当初制定HTTP的初衷,这样有助于我们更好地理解。

1.2.1 为知识共享而规划Web

1989年3月,互联网还只属于少数人。在这一互联网的黎明期,HTTP诞生了。



003

① HTTP通常被译为超文本传输协议,但这种译法并不严谨。严谨的译名应该为“超文本转移协议”。但是前一译法已约定俗成,本书将会沿用。有兴趣的读者可参考图灵社区的相关讨论:<http://www.ituring.com.cn/article/1817>。

——译者注