

TURING

图灵程序设计丛书

*High Performance
Browser Networking*



Web性能

权威指南

[加] Ilya Grigorik 著
李松峰 译

O'REILLY®

人民邮电出版社
POSTS & TELECOM PRESS



图灵程序设计丛书

Web性能权威指南

High Performance Browser Networking

[加] Ilya Grigorik 著
李松峰 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'Reilly Media, Inc. 授权人民邮电出版社出版

人民邮电出版社
北 京

图书在版编目 (C I P) 数据

Web性能权威指南 / (加) 格里高利克
(Grigorik, I.) 著 ; 李松峰译. — 北京 : 人民邮电出
版社, 2014. 5

(图灵程序设计丛书)

书名原文: High performance browser networking
ISBN 978-7-115-34910-1

I. ①W… II. ①格… ②李… III. ①计算机网络—程
序设计 IV. ①TP393.09

中国版本图书馆CIP数据核字(2014)第043538号

内 容 提 要

本书是谷歌公司高性能团队核心成员的权威之作, 堪称实战经验与规范解读完美结合的产物。本书目标是涵盖 Web 开发者技术体系中应该掌握的所有网络及性能优化知识。全书以性能优化为主线, 从 TCP、UDP 和 TLS 协议讲起, 解释了如何针对这几种协议和基础设施来优化应用。然后深入探讨了无线和移动网络的工作机制。最后, 揭示了 HTTP 协议的底层细节, 同时详细介绍了 HTTP 2.0、XHR、SSE、WebSocket、WebRTC 和 DataChannel 等现代浏览器新增的具有革命性的新能力。

本书适合所有 Web 应用及站点开发人员阅读, 包括但不限于前端、后端、运维、大数据分析、UI/UX、存储、视频、实时消息, 以及性能工程师。

◆ 著 [加] Ilya Grigorik

译 李松峰

责任编辑 李 瑛

执行编辑 李 静

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>

北京鑫正大印刷有限公司印刷

◆ 开本: 800×1000 1/16

印张: 21

字数: 400千字

2014年5月第1版

印数: 1~5 000册

2014年5月北京第1次印刷

著作权合同登记号 图字: 01-2013-7651号

定价: 69.00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

©2013 by Ilya Grigorik.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2014. Authorized translation of the English edition, 2013 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

英文原版由 O'Reilly Media, Inc. 出版 2013。

简体中文版由人民邮电出版社出版，2014。英文原版的翻译得到 O'Reilly Media, Inc. 的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly Media, Inc. 的许可。

版权所有，未得书面许可，本书的任何部分和全部不得以任何形式重制。

O'Reilly Media, Inc.介绍

O'Reilly Media 通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自 1978 年开始，O'Reilly 一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly 的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly 为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了 Make 杂志，从而成为 DIY 革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly 的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新兴产业的革命性思想。作为技术人士获取信息的选择，O'Reilly 现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项 O'Reilly 的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar 博客有口皆碑。”

——Wired

“O'Reilly 凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference 是聚集关键思想领袖的绝对典范。”

——CRN

“一本 O'Reilly 的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim 是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照 Yogi Berra 的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去 Tim 似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

Steve Souders推荐序

“合格的开发者知道怎么做，而优秀的开发者知道为什么那么做。”

相信每一位读者看完这句话，一定打心眼儿里赞同。我们都希望自己能够理解身边的各种系统，同时还能跟别人讲得明白。然而，如果你是一名 Web 开发者，那很可能距离这个目标会越来越远。

Web 开发的分工越来越细。你在做哪一类 Web 开发？前端？后端？运维？大数据分析？UI/UX？存储？视频？实时消息？我还想再加上一个角色——性能工程师。

钻研基础知识与紧跟最新动向本身是一对矛盾，很难平衡。可是，没有基础，那只能是“墙上芦苇，头重脚轻根底浅”。光知道表面上的那点东西可不行。需要解决难题时，发生异常状况时，理解基础知识的人会脱颖而出。

正因为如此，我才说这本书非常重要，不能不看。如果你搞的是 Web 开发，那你技术体系的根基就是 Web 和它赖以存在的大量网络协议：TCP、TLS、UDP、HTTP，等等。这些协议分别有各自的性能特点和优化技巧，为开发高性能应用，你必须理解为什么网络那么运行。

说实话，我真为想读这本书的你感到庆幸！要是我刚刚接触 Web 编程时有这样一本书就好了。那样，就会有一位真正理解网络的人为我释疑解惑，告诉我那些标准和规范的要点，填充我技术体系中的空白。这本书的作者 Ilya Grigorik，是少见的网络编程专家，而本书堪称实战经验与规范解读完美结合的产物。

本书中，作者解释了网络编程中的很多为什么：为什么延迟是性能瓶颈？为什么 TCP 并不总是最优传输机制，而 UDP 有时候反而是更好的选择？为什么重用连接是关键性的优化策略？然后，他又更进一步，给出改进网络性能的具体建议。想要降低延迟？在靠近客户端的服务器上完成会话。想要提高连接重用率？保持连接持

久化。正是这种提出问题、分析问题和解决问题的模式，让本书内容极为贴近实战，接地气。

除了全面探讨网络的基础知识，作者还详细讲解了协议和浏览器的最新进展。讲了 HTTP 2.0 的诸多优点，回顾了 XHR 及其催生 CORS (Cross-Origin Resource Sharing, 跨源资源共享) 的局限性，还有 SSE (Server-Sent Events, 服务器发送事件)、WebSockets 和 WebRTC。让我们彻底跟上了浏览器网络技术栈的最新进展。

从性能角度分析，基础和最新进展是本书特色，也是本书贯穿始终的主线。正是性能这个视角，让我们理解了网络开发中的那么多为什么，明白了这些东西怎么影响我们的网站和用户。本书把抽象的规范变成了可操作的建议，让我们可以马上学以致用去优化网站，去创造最佳用户体验。这才是最重要的。所以，一定不要错过这本书！

——Steve Souders

世界级 Web 性能专家、谷歌公司高性能工程师
《高性能网站建设指南》等畅销书作者，2013¹

注 1：2014 年 3 月，Steve Souders 加入国际知名云加速平台 Fastly，任首席性能官。——编者注

前言

对所有程序员而言，Web 浏览器是今天部署最为广泛的应用平台。每一部智能手机，每一台平板、笔记本和桌面电脑，以及介于这些之间的其他设备，都安装了 Web 浏览器。据预测，到 2020 年，能上网的设备将突破 200 亿部，其中每一部都会安装浏览器，而且最低限度都会支持 Wi-Fi 或蜂窝连接。至于运行的是什么平台，设备是谁制造的，操作系统是哪个版本，都无所谓，反正每部设备都会带一个浏览器。就是这个浏览器，它的功能正变得越来越强大。

回首从前，那时的浏览器与现在我们使用的浏览器完全不能同日而语。浏览器革命还是最近几年的事：HTML 和 CSS 构成表现层，JavaScript 则成为 Web 上的新“汇编”语言，而新的 HTML5 API 仍在不断改进，致力于为交付吸引人的高性能应用提供新功能。可以说，有史以来还没有哪项技术或平台，在部署或装机率方面能和今天的浏览器相提并论。所以，这里有无限的机会，创新也无处不在。

而且，浏览器中网络基础设施的快速发展与创新，同样是天下无敌。过去，我们能实现的交互仅限于简单的 HTTP 请求和响应；如今，高效的流式传输、双向实时通信，以及交付自定义协议，甚至端到端视频会议和两端之间直接交换数据，都已经成为现实。而且，所有这些只不过是几十行 JavaScript 代码的事儿。

然后呢？几十亿设备互联，已有和新在线服务吸引的用户越来越多，只有高性能 Web 应用才谈得上竞争力。速度是关键！事实上，对某些应用来说，速度决定命运。要开发出高性能的 Web 应用，必须透彻理解浏览器及其网络交互机制，而这正是本书的主题。

关于本书

本书目标是涵盖开发者应该掌握的所有网络知识：网络开发中要用到哪些协议，这些协议有什么固有的局限性，如何针对底层网络优化自己的应用，浏览器提供了哪些网

络相关的功能，以及什么时候需要用到它们。

我们将从 TCP、UDP 和 TLS 协议的内部工作原理讲起，向大家解释如何针对这几种协议和基础设施来优化我们的应用。然后深入地探讨无线和移动网络的工作机制，以无线电为媒介的通信可大不一样。对此，我们将围绕如何设计和架构应用，讨论它们各自的痛点所在。最后，我们再揭示 HTTP 协议的底层细节，同时详细介绍浏览器新增的一些令人激动的能力：

- 即将到来的 HTTP 2.0 的诸多改进；
- XHR 的新特性和新功能；
- 通过 SSE 发送数据流；
- 通过 WebSocket 实现双向通信；
- 通过 WebRTC 实现端到端的音频和视频通信；
- 通过 DataChannel 实现端到端的数据交换。

要设计和开发高性能的应用，必须理解每一位数据是如何交付的，必须理解每一种传输机制和相关协议的特点。毕竟，等待网络是我们应用最大的性能瓶颈，再怎么优化渲染 JavaScript 或其他方面，也抵不上网络优化！本书的目标就是告诉读者怎么消除等待时间，利用现有网络实现最大的性能优化。

本书全面介绍了 Web 性能优化的知识和技术，适合对构建和交付高性能应用感兴趣的所有读者。简单地说，如果你不满足于那些枯燥的检查表，而更希望知晓浏览器乃至底层协议的真实工作过程，就应该读一读这本书。本书既会对配置和架构给出实用建议，也会探讨为达成优化目标而必须考虑的因素和权衡的要点，既讲“怎么办”，也讲“为什么”。



本书重点讨论与浏览器应用相关的各种协议及特性。不过，关于 TCP、UDP、TLS、HTTP，乃至其他每一种协议的讨论，同样也适用于本地应用，而且不局限于任何平台。

排版约定

本书使用的排版约定如下。

- 楷体
表示新的术语。

- 等宽字体

表示程序片段，也用于在正文中表示程序中使用的变量、函数名、命令行代码、环境变量、语句和关键词等代码文本。

- 加粗的等宽字体

表示应该由用户逐字输入的命令或者其他文本。

- 倾斜的等宽字体

表示应该由用户输入的值或根据上下文决定的值替换的文本。



这个图标代表小窍门、建议或说明。



这个图标代表警告信息。

Safari® Books Online



Safari Books Online (www.safaribooksonline.com) 是应需而变的数字图书馆。它同时以图书和视频的形式出版世界顶级技术和商务作家的专业作品。

Safari Books Online 是技术专家、软件开发人员、Web 设计师、商务人士和创意人士开展调研、解决问题、学习和认证培训的第一手资料。

对于组织团体、政府机构和个人，Safari Books Online 提供各种产品组合和灵活的定价策略。用户可通过一个功能完备的数据库检索系统访问 O'Reilly Media、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、Focal Press、Cisco Press、John Wiley & Sons、Syngress、Morgan Kaufmann、IBM Redbooks、Packt、Adobe Press、FT Press、Apress、Manning、New Riders、McGraw-Hill、Jones & Bartlett、Course Technology 以及其他几十家出版社的上千种图书、培训视频和正式出版之前的书稿。要了解 Safari Books Online 的更多信息，我们网上见。

联系我们

请把对本书的评价和问题发给出版社。

美国：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472

中国：

北京市西城区西直门南大街 2 号成铭大厦 C 座 807 室 (100035)
奥莱利技术咨询 (北京) 有限公司

O'Reilly 的每一本书都有专属网页，你可以在那儿找到本书的相关信息，包括勘误表、示例代码以及其他信息。本书的网站地址是：

<http://shop.oreilly.com/product/0636920028084.do>

对于本书的评论和技术性问题，请发送电子邮件到：

bookquestions@oreilly.com

要了解更多 O'Reilly 图书、培训课程、会议和新闻的信息，请访问以下网站：

<http://www.oreilly.com>

我们在 Facebook 的地址如下：

<http://facebook.com/oreilly>

请关注我们的 Twitter 动态：

<http://twitter.com/oreillymedia>

我们的 YouTube 视频地址如下：

<http://www.youtube.com/oreillymedia>

目录

Steve Souders 推荐序	XIII
-------------------	------

前言	XV
----	----

第一部分 网络技术概览

第 1 章 延迟与带宽	3
1.1 速度是关键	3
1.2 延迟的构成	4
1.3 光速与传播延迟	6
1.4 延迟的最后一公里	7
1.5 网络核心的带宽	8
1.6 网络边缘的带宽	9
1.7 目标：高带宽和低延迟	10
第 2 章 TCP 的构成	13
2.1 三次握手	14
2.2 拥塞预防及控制	16
2.2.1 流量控制	16
2.2.2 慢启动	18
2.2.3 拥塞预防	24
2.3 带宽延迟积	25
2.4 队首阻塞	27
2.5 针对 TCP 的优化建议	28
2.5.1 服务器配置调优	29
2.5.2 应用程序行为调优	30

2.5.3 性能检查清单	30
第3章 UDP 的构成	31
3.1 无协议服务	32
3.2 UDP 与网络地址转换器	34
3.2.1 连接状态超时	35
3.2.2 NAT 穿透	36
3.2.3 STUN、TURN 与 ICE	37
3.3 针对 UDP 的优化建议	39
第4章 传输层安全 (TLS)	41
4.1 加密、身份验证与完整性	42
4.2 TLS 握手	44
4.2.1 应用层协议协商 (ALPN)	46
4.2.2 服务器名称指示 (SNI)	47
4.3 TLS 会话恢复	48
4.3.1 会话标识符	48
4.3.2 会话记录单	49
4.4 信任链与证书颁发机构	50
4.5 证书撤销	52
4.5.1 证书撤销名单 (CRL)	53
4.5.2 在线证书状态协议 (OCSP)	54
4.6 TLS 记录协议	54
4.7 针对 TLS 的优化建议	55
4.7.1 计算成本	55
4.7.2 尽早完成 (握手)	56
4.7.3 会话缓存与无状态恢复	58
4.7.4 TLS 记录大小	59
4.7.5 TLS 压缩	60
4.7.6 证书链的长度	61
4.7.7 OCSP 封套	62
4.7.8 HTTP 严格传输安全 (HSTS)	62
4.8 性能检查清单	63
4.9 测试与验证	64

第二部分 无线网络性能

第5章 无线网络概览	69
-------------------------	-----------

5.1	无所不在的连接	69
5.2	无线网络的类型	70
5.3	无线网络的性能基础	71
5.3.1	带宽	71
5.3.2	信号强度	74
5.3.3	调制	75
5.4	测量现实中的无线性能	76
第6章	Wi-Fi	79
6.1	从以太网到无线局域网	79
6.2	Wi-Fi 标准及功能	81
6.3	测量和优化 Wi-Fi 性能	81
6.4	针对 Wi-Fi 的优化建议	84
6.4.1	利用不计流量的带宽	84
6.4.2	适应可变带宽	85
6.4.3	适应可变的延迟时间	86
第7章	移动网络	87
7.1	G 字号移动网络简介	87
7.1.1	最早提供数据服务的 2G	88
7.1.2	3GPP 与 3GPP2	89
7.1.3	3G 技术的演进	91
7.1.4	IMT-Advanced 的 4G 要求	93
7.1.5	长期演进 (LTE)	94
7.1.6	HSPA+ 推进世界范围内的 4G 普及	95
7.1.7	为多代并存的未来规划	96
7.2	设备特性及能力	97
7.3	无线电资源控制器 (RRC)	99
7.3.1	3G、4G 和 Wi-Fi 对电源的要求	101
7.3.2	LTE RRC 状态机	102
7.3.3	HSPA 与 HSPA+ (UMTS) RRC 状态机	104
7.3.4	EV-DO (CDMA) RRC 状态机	106
7.3.5	低效率的周期性传输	107
7.4	端到端的运营商架构	108
7.4.1	无线接入网络 (RAN)	108
7.4.2	核心网络	110
7.4.3	回程容量与延迟	112
7.5	移动网络中的分组流	113
7.5.1	初始化请求	113

7.5.2 入站数据流	116
7.6 异质网络 (HetNet)	117
7.7 真实的 3G、4G 和 Wi-Fi 性能	119
第 8 章 移动网络的优化建议	121
8.1 节约用电	122
8.2 消除周期性及无效的数据传输	124
8.3 预测网络延迟上限	126
8.3.1 考虑 RRC 状态切换	127
8.3.2 解耦用户交互与网络通信	128
8.4 面对多网络接口并存的现实	128
8.5 爆发传输数据并转为空闲	130
8.6 把负载转移到 Wi-Fi 网络	131
8.7 遵从协议和应用最佳实践	131

第三部分 HTTP

第 9 章 HTTP 简史	135
9.1 HTTP 0.9: 只有一行的协议	135
9.2 HTTP 1.0: 迅速发展及参考性 RFC	136
9.3 HTTP 1.1: 互联网标准	138
9.4 HTTP 2.0: 改进传输性能	141
第 10 章 Web 性能要点	143
10.1 超文本、网页和 Web 应用	144
10.2 剖析现代 Web 应用	146
10.2.1 速度、性能与用户期望	147
10.2.2 分析资源瀑布	148
10.3 性能来源: 计算、渲染和网络访问	151
10.3.1 更多带宽其实不 (太) 重要	152
10.3.2 延迟是性能瓶颈	152
10.4 人造和真实用户性能度量	154
10.5 针对浏览器的优化建议	157
第 11 章 HTTP 1.x	161
11.1 持久连接的优点	163
11.2 HTTP 管道	165
11.3 使用多个 TCP 连接	169

11.4	域名分区	171
11.5	度量和控制协议开销	173
11.6	连接与拼合	174
11.7	嵌入资源	177
第 12 章	HTTP 2.0	179
12.1	历史及其与 SPDY 的渊源	180
12.2	走向 HTTP 2.0	181
12.3	设计和技术目标	182
12.3.1	二进制分帧层	183
12.3.2	流、消息和帧	184
12.3.3	多向请求与响应	185
12.3.4	请求优先级	186
12.3.5	每个来源一个连接	188
12.3.6	流量控制	189
12.3.7	服务器推送	190
12.3.8	首部压缩	192
12.3.9	有效的 HTTP 2.0 升级与发现	194
12.4	二进制分帧简介	196
12.4.1	发起新流	197
12.4.2	发送应用数据	198
12.4.3	HTTP 2.0 帧数据流分析	199
第 13 章	优化应用的交付	201
13.1	经典的性能优化最佳实践	203
13.1.1	在客户端缓存资源	204
13.1.2	压缩传输的数据	205
13.1.3	消除不必要的请求字节	206
13.1.4	并行处理请求和响应	207
13.2	针对 HTTP 1.x 的优化建议	208
13.3	针对 HTTP 2.0 的优化建议	209
13.3.1	去掉对 1.x 的优化	209
13.3.2	双协议应用策略	210
13.3.3	1.x 与 2.0 的相互转换	212
13.3.4	评估服务器质量与性能	213
13.3.5	2.0 与 TLS	214
13.3.6	负载均衡器、代理及应用服务器	215

第四部分 浏览器 API 与协议

第 14 章 浏览器网络概述	219
14.1 连接管理与优化	220
14.2 网络安全与沙箱	222
14.3 资源与客户端状态缓存	222
14.4 应用 API 与协议	223
第 15 章 XMLHttpRequest	225
15.1 XHR 简史	226
15.2 跨源资源共享 (CORS)	227
15.3 通过 XHR 下载数据	230
15.4 通过 XHR 上传数据	231
15.5 监控下载和上传进度	233
15.6 通过 XHR 实现流式数据传输	234
15.7 实时通知与交付	236
15.7.1 通过 XHR 实现轮询	237
15.7.2 通过 XHR 实现长轮询	238
15.8 XHR 使用场景及性能	240
第 16 章 服务器发送事件	243
16.1 EventSource API	243
16.2 Event Stream 协议	245
16.3 SSE 使用场景及性能	248
第 17 章 WebSocket	251
17.1 WebSocket API	252
17.1.1 WS 与 WSS	253
17.1.2 接收文本和二进制数据	253
17.1.3 发送文本和二进制数据	255
17.1.4 子协议协商	256
17.2 WebSocket 协议	257
17.2.1 二进制分帧层	258
17.2.2 协议扩展	260
17.2.3 HTTP 升级协商	261
17.3 WebSocket 使用场景及性能	264
17.3.1 请求和响应流	264
17.3.2 消息开销	265