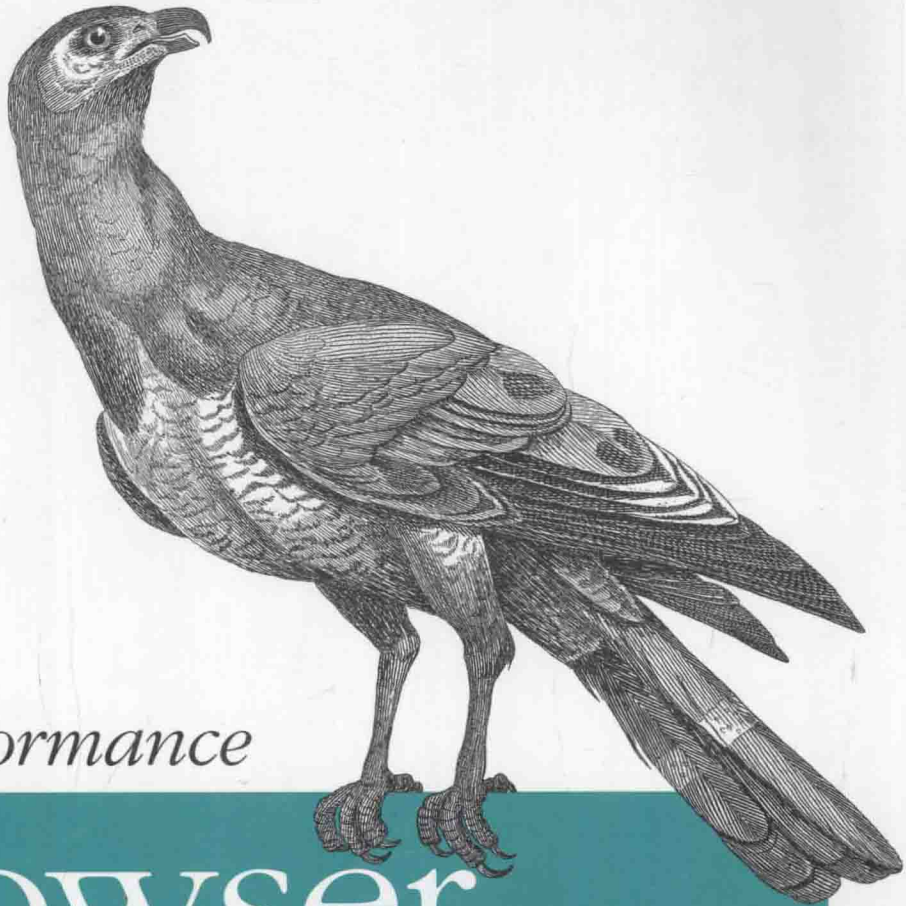
高性能浏览器网络（影印版）



*High Performance*

# Browser Networking

O'REILLY®

東南大學出版社

*Ilya Grigorik* 著

# 高性能浏览器网络（影印版）

# High Performance Browser Networking

*Ilya Grigorik* 著

O'REILLY®

Beijing · Cambridge · Farnham · Köln · Sebastopol · Tokyo

O'Reilly Media, Inc.授权东南大学出版社出版

南京 东南大学出版社

**图书在版编目（CIP）数据**

高性能浏览器网络:英文/(美)格里戈利克(Grigorik, I.)著.—影印本.—南京：东南大学出版社，2014.9

书名原文：High Performance Browser Networking

ISBN 978-7-5641-4961-1

I. ① 高… II. ①格… III. ①浏览器－英文 IV. ① TP393.092

中国版本图书馆 CIP 数据核字（2014）第 102959 号

江苏省版权局著作权合同登记

图字：10-2014-146 号

©2013 by O'Reilly Media, Inc.

Reprint of the English Edition, jointly published by O'Reilly Media, Inc. and Southeast University Press, 2014. Authorized reprint of the original English edition, 2014 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。

英文影印版由东南大学出版社出版 2014。此影印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。

高性能浏览器网络（影印版）

出版发行：东南大学出版社

地　　址：南京四牌楼 2 号　　邮编：210096

出 版 人：江建中

网　　址：http://www.seupress.com

电子邮件：press@seupress.com

印　　刷：常州市武进第三印刷有限公司

开　　本：787 毫米 × 980 毫米　16 开本

印　　张：25.25

字　　数：494 千字

版　　次：2014 年 9 月第 1 版

印　　次：2014 年 9 月第 1 次印刷

书　　号：ISBN 978-7-5641-4961-1

定　　价：74.00 元

本社图书若有印装质量问题，请直接与营销部联系。电话（传真）：025-83791830

# Foreword

"Good developers know how things work. Great developers know why things work."

We all resonate with this adage. We want to be that person who understands and can explain the underpinning of the systems we depend on. And yet, if you're a web developer, you might be moving in the opposite direction.

Web development is becoming more and more specialized. What kind of web developer are you? Frontend? Backend? Ops? Big data analytics? UI/UX? Storage? Video? Messaging? I would add "Performance Engineer" making that list of possible specializations even longer.

It's hard to balance studying the foundations of the technology stack with the need to keep up with the latest innovations. And yet, if we don't understand the foundation our knowledge is hollow, shallow. Knowing how to use the topmost layers of the technology stack isn't enough. When the complex problems need to be solved, when the inexplicable happens, the person who understands the foundation leads the way.

That's why *High Performance Browser Networking* is an important book. If you're a web developer, the foundation of your technology stack is the Web and the myriad of networking protocols it rides on: TCP, TLS, UDP, HTTP, and many others. Each of these protocols has its own performance characteristics and optimizations, and to build high performance applications you need to understand why the network behaves the way it does.

Thank goodness you've found your way to this book. I wish I had this book when I started web programming. I was able to move forward by listening to people who understood the why of networking and read specifications to fill in the gaps. High Performance Browser Networking combines the expertise of a networking guru, Ilya Grigorik, with the necessary information from the many relevant specifications, all woven together in one place.

In *High Performance Browser Networking*, Ilya explains many whys of networking: Why latency is the performance bottleneck. Why TCP isn't always the best transport mechanism and UDP might be your better choice. Why reusing connections is a critical optimization. He then goes even further by providing specific actions for improving networking performance. Want to reduce latency? Terminate sessions at a server closer to the client. Want to increase connection reuse? Enable connection keep-alive. The combination of understanding what to do and why it matters turns this knowledge into action.

Ilya explains the foundation of networking and builds on that to introduce the latest advances in protocols and browsers. The benefits of HTTP 2.0 are explained. XHR is reviewed and its limitations motivate the introduction of Cross-Origin Resource Sharing. Server-Sent Events, WebSockets, and WebRTC are also covered, bringing us up to date on the latest in browser networking.

Viewing the foundation and latest advances in networking from the perspective of performance is what ties the book together. Performance is the context that helps us see the why of networking and translate that into how it affects our website and our users. It transforms abstract specifications into tools that we can wield to optimize our websites and create the best user experience possible. That's important. That's why you should read this book.

—Steve Souders, Head Performance Engineer, Google, 2013



此为试读，需要完整PDF请访问：www.ertongbook.com

# Preface

The web browser is the most widespread deployment platform available to developers today: it is installed on every smartphone, tablet, laptop, desktop, and every other form factor in between. In fact, current cumulative industry growth projections put us on track for 20 billion connected devices by 2020—each with a browser, and at the very least, WiFi or a cellular connection. The type of platform, manufacturer of the device, or the version of the operating system do not matter—each and every device will have a web browser, which by itself is getting more feature rich each day.

The browser of yesterday looks nothing like what we now have access to, thanks to all the recent innovations: HTML and CSS form the presentation layer, JavaScript is the new assembly language of the Web, and new HTML5 APIs are continuing to improve and expose new platform capabilities for delivering engaging, high-performance applications. There is simply no other technology, or platform, that has ever had the reach or the distribution that is made available to us today when we develop for the browser. And where there is big opportunity, innovation always follows.

In fact, there is no better example of the rapid progress and innovation than the networking infrastructure within the browser. Historically, we have been restricted to simple HTTP request-response interactions, and today we have mechanisms for efficient streaming, bidirectional and real-time communication, ability to deliver custom application protocols, and even peer-to-peer videoconferencing and data delivery directly between the peers—all with a few dozen lines of JavaScript.

The net result? Billions of connected devices, a swelling userbase for existing and new online services, and high demand for high-performance web applications. Speed is a feature, and in fact, for some applications it is *the feature*, and delivering a high-performance web application requires a solid foundation in how the browser and the network interact. That is the subject of this book.

## About This Book

Our goal is to cover what every developer should know about the network: what protocols are being used and their inherent limitations, how to best optimize your applications for the underlying network, and what networking capabilities the browser offers and when to use them.

In the process, we will look at the internals of TCP, UDP, and TLS protocols, and how to optimize our applications and infrastructure for each one. Then we'll take a deep dive into how the wireless and mobile networks work under the hood—this radio thing, it's very different—and discuss its implications for how we design and architect our applications. Finally, we will dissect how the HTTP protocol works under the hood and investigate the many new and exciting networking capabilities in the browser:

- Upcoming HTTP 2.0 improvements
- New XHR features and capabilities
- Data streaming with Server-Sent Events
- Bidirectional communication with WebSocket
- Peer-to-peer video and audio communication with WebRTC
- Peer-to-peer data exchange with DataChannel

Understanding how the individual bits are delivered, and the properties of each transport and protocol in use are essential knowledge for delivering high-performance applications. After all, if our applications are blocked waiting on the network, then no amount of rendering, JavaScript, or any other form of optimization will help! Our goal is to eliminate this wait time by getting the best possible performance from the network.

*High-Performance Browser Networking* will be of interest to anyone interested in optimizing the delivery and performance of her applications, and more generally, curious minds that are not satisfied with a simple checklist but want to know how the browser and the underlying protocols actually work under the hood. The "how" and the "why" go hand in hand: we'll cover practical advice about configuration and architecture, and we'll also explore the trade-offs and the underlying reasons for each optimization.

Our primary focus is on the protocols and their properties with respect to applications running in the browser. However, all the discussions on TCP, UDP, TLS, HTTP, and just about every other protocol we will cover are also directly applicable to native applications, regardless of the platform.

## Conventions Used in This Book

The following typographical conventions are used in this book:

*Italic*
: Indicates new terms, URLs, email addresses, filenames, and file extensions.

`Constant width`
: Used for program listings, as well as within paragraphs to refer to program elements such as variable or function names, databases, data types, environment variables, statements, and keywords.

**`Constant width bold`**
: Shows commands or other text that should be typed literally by the user.

*`Constant width italic`*
: Shows text that should be replaced with user-supplied values or by values determined by context.

This icon signifies a tip, suggestion, or general note.

This icon indicates a warning or caution.

## Safari® Books Online

*Safari Books Online* is an on-demand digital library that delivers expert content in both book and video form from the world's leading authors in technology and business.

Technology professionals, software developers, web designers, and business and creative professionals use Safari Books Online as their primary resource for research, problem solving, learning, and certification training.

Safari Books Online offers a range of product mixes and pricing programs for organizations, government agencies, and individuals. Subscribers have access to thousands of books, training videos, and prepublication manuscripts in one fully searchable database from publishers like 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, and dozens more. For more information about Safari Books Online, please visit us online.

## How to Contact Us

Please address comments and questions concerning this book to the publisher:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at *http://oreil.ly/high-performance-browser*.

To comment or ask technical questions about this book, send email to *bookquestions@oreilly.com*.

For more information about our books, courses, conferences, and news, see our website at *http://www.oreilly.com*.

Find us on Facebook: *http://facebook.com/oreilly*

Follow us on Twitter: *http://twitter.com/oreillymedia*

Watch us on YouTube: *http://www.youtube.com/oreillymedia*

# Table of Contents

**Foreword** ..... xi
**Preface** ..... xiii

## Part I. Networking 101

**1. Primer on Latency and Bandwidth** ..... 3
Speed Is a Feature 3
The Many Components of Latency 4
Speed of Light and Propagation Latency 6
Last-Mile Latency 8
Bandwidth in Core Networks 9
Bandwidth at the Network Edge 10
Delivering Higher Bandwidth and Lower Latencies 11

**2. Building Blocks of TCP** ..... 13
Three-Way Handshake 14
Congestion Avoidance and Control 16
Flow Control 17
Slow-Start 19
Congestion Avoidance 26
Bandwidth-Delay Product 28
Head-of-Line Blocking 30
Optimizing for TCP 32
Tuning Server Configuration 32
Tuning Application Behavior 34
Performance Checklist 34

**3. Building Blocks of UDP** ..... 35
Null Protocol Services 36

UDP and Network Address Translators 38
Connection-State Timeouts 39
NAT Traversal 40
STUN, TURN, and ICE 41
Optimizing for UDP 44

**4. Transport Layer Security (TLS) ..... 47**
Encryption, Authentication, and Integrity 48
TLS Handshake 50
Application Layer Protocol Negotiation (ALPN) 53
Server Name Indication (SNI) 54
TLS Session Resumption 55
Session Identifiers 55
Session Tickets 57
Chain of Trust and Certificate Authorities 57
Certificate Revocation 61
Certificate Revocation List (CRL) 61
Online Certificate Status Protocol (OCSP) 62
TLS Record Protocol 62
Optimizing for TLS 63
Computational Costs 64
Early Termination 65
Session Caching and Stateless Resumption 67
TLS Record Size 68
TLS Compression 69
Certificate-Chain Length 70
OCSP Stapling 71
HTTP Strict Transport Security (HSTS) 72
Performance Checklist 73
Testing and Verification 73

**Part II. Performance of Wireless Networks**

**5. Introduction to Wireless Networks ..... 79**
Ubiquitous Connectivity 79
Types of Wireless Networks 80
Performance Fundamentals of Wireless Networks 81
Bandwidth 82
Signal Power 85
Modulation 86



此为试读，需要完整PDF请访问：www.ertongbook.com

Measuring Real-World Wireless Performance 87

**6. WiFi** ..... **89**
From Ethernet to a Wireless LAN 89
WiFi Standards and Features 91
Measuring and Optimizing WiFi Performance 92
Packet Loss in WiFi Networks 94
Optimizing for WiFi Networks 95
Leverage Unmetered Bandwidth 95
Adapt to Variable Bandwidth 96
Adapt to Variable Latency 97

**7. Mobile Networks** ..... **99**
Brief History of the G's 99
First Data Services with 2G 100
3GPP and 3GPP2 Partnerships 101
Evolution of 3G Technologies 103
IMT-Advanced 4G Requirements 105
Long Term Evolution (LTE) 106
HSPA+ is Leading Worldwide 4G Adoption 107
Building for the Multigeneration Future 109
Device Features and Capabilities 111
User Equipment Category 111
Radio Resource Controller (RRC) 113
3G, 4G, and WiFi Power Requirements 115
LTE RRC State Machine 116
HSPA and HSPA+ (UMTS) RRC State Machine 119
EV-DO (CDMA) RRC State Machine 120
Inefficiency of Periodic Transfers 121
End-to-End Carrier Architecture 123
Radio Access Network (RAN) 123
Core Network (CN) 125
Backhaul Capacity and Latency 128
Packet Flow in a Mobile Network 129
Initiating a Request 129
Inbound Data Flow 132
Heterogeneous Networks (HetNets) 133
Real-World 3G, 4G, and WiFi Performance 135

**8. Optimizing for Mobile Networks** ..... **139**
Preserve Battery Power 140
Eliminate Periodic and Inefficient Data Transfers 142

Eliminate Unnecessary Application Keepalives 144
Anticipate Network Latency Overhead 145
Account for RRC State Transitions 146
Decouple User Interactions from Network Communication 146
Design for Variable Network Interface Availability 147
Burst Your Data and Return to Idle 149
Offload to WiFi Networks 150
Apply Protocol and Application Best Practices 150

**Part III. HTTP**

**9. Brief History of HTTP ... 155**
HTTP 0.9: The One-Line Protocol 155
HTTP 1.0: Rapid Growth and Informational RFC 157
HTTP 1.1: Internet Standard 159
HTTP 2.0: Improving Transport Performance 161

**10. Primer on Web Performance ... 165**
Hypertext, Web Pages, and Web Applications 166
Anatomy of a Modern Web Application 168
Speed, Performance, and Human Perception 170
Analyzing the Resource Waterfall 171
Performance Pillars: Computing, Rendering, Networking 176
More Bandwidth Doesn't Matter (Much) 176
Latency as a Performance Bottleneck 177
Synthetic and Real-User Performance Measurement 179
Browser Optimization 183

**11. HTTP 1.X ... 187**
Benefits of Keepalive Connections 189
HTTP Pipelining 192
Using Multiple TCP Connections 196
Domain Sharding 198
Measuring and Controlling Protocol Overhead 200
Concatenation and Spriting 201
Resource Inlining 204

**12. HTTP 2.0 ... 207**
History and Relationship to SPDY 208
The Road to HTTP 2.0 209
Design and Technical Goals 211

Binary Framing Layer 211
Streams, Messages, and Frames 212
Request and Response Multiplexing 214
Request Prioritization 215
One Connection Per Origin 217
Flow Control 218
Server Push 219
Header Compression 222
Efficient HTTP 2.0 Upgrade and Discovery 224
Brief Introduction to Binary Framing 226
Initiating a New Stream 229
Sending Application Data 230
Analyzing HTTP 2.0 Frame Data Flow 230

**13. Optimizing Application Delivery..... 233**
Evergreen Performance Best Practices 235
Cache Resources on the Client 236
Compress Transferred Data 237
Eliminate Unnecessary Request Bytes 238
Parallelize Request and Response Processing 239
Optimizing for HTTP 1.x 241
Optimizing for HTTP 2.0 241
Removing 1.x Optimizations 242
Dual-Protocol Application Strategies 244
Translating 1.x to 2.0 and Back 245
Evaluating Server Quality and Performance 247
Speaking 2.0 with and without TLS 247
Load Balancers, Proxies, and Application Servers 248

**Part IV. Browser APIs and Protocols**

**14. Primer on Browser Networking..... 253**
Connection Management and Optimization 254
Network Security and Sandboxing 256
Resource and Client State Caching 257
Application APIs and Protocols 258

**15. XMLHttpRequest..... 261**
Brief History of XHR 262
Cross-Origin Resource Sharing (CORS) 263
Downloading Data with XHR 266

Uploading Data with XHR 268
Monitoring Download and Upload Progress 269
Streaming Data with XHR 271
Real-Time Notifications and Delivery 273
Polling with XHR 274
Long-Polling with XHR 276
XHR Use Cases and Performance 278

**16. Server-Sent Events (SSE) ..... 279**
EventSource API 279
Event Stream Protocol 282
SSE Use Cases and Performance 285

**17. WebSocket ..... 287**
WebSocket API 288
WS and WSS URL Schemes 289
Receiving Text and Binary Data 290
Sending Text and Binary Data 291
Subprotocol Negotiation 293
WebSocket Protocol 294
Binary Framing Layer 295
Protocol Extensions 297
HTTP Upgrade Negotiation 298
WebSocket Use Cases and Performance 301
Request and Response Streaming 302
Message Overhead 303
Data Efficiency and Compression 304
Custom Application Protocols 304
Deploying WebSocket Infrastructure 305
Performance Checklist 307

**18. WebRTC ..... 309**
Standards and Development of WebRTC 310
Audio and Video Engines 311
Acquiring Audio and Video with getUserMedia 312
Real-Time Network Transports 315
Brief Introduction to RTCPeerConnection API 317
Establishing a Peer-to-Peer Connection 319
Signaling and Session Negotiation 320
Session Description Protocol (SDP) 322
Interactive Connectivity Establishment (ICE) 325
Incremental Provisioning (Trickle ICE) 328

Tracking ICE Gathering and Connectivity Status 329
Putting It All Together 332
Delivering Media and Application Data 337
Secure Communication with DTLS 337
Delivering Media with SRTP and SRTCP 340
Delivering application data with SCTP 344
DataChannel 348
Setup and Negotiation 350
Configuring Message Order and Reliability 353
Partially Reliable Delivery and Message Size 355
WebRTC Use Cases and Performance 356
Audio, Video, and Data Streaming 356
Multiparty Architectures 358
Infrastructure and Capacity Planning 359
Data Efficiency and Compression 361
Performance Checklist 361

**Index** ..... **363**

# PART I
# Networking 101

此为试读,需要完整PDF请访问: www.ertongbook.com