

# The End-to-End Performance Imperative

---

Delivering fast, responsive, and available Internet-facing applications has become more critical and challenging. Learn a compelling strategy to tackle the most common problems facing leading “always-on” businesses.

## Primary Trends Affecting Your Application Performance

There are three reasons why your web and mobile application performance matters even more than before:

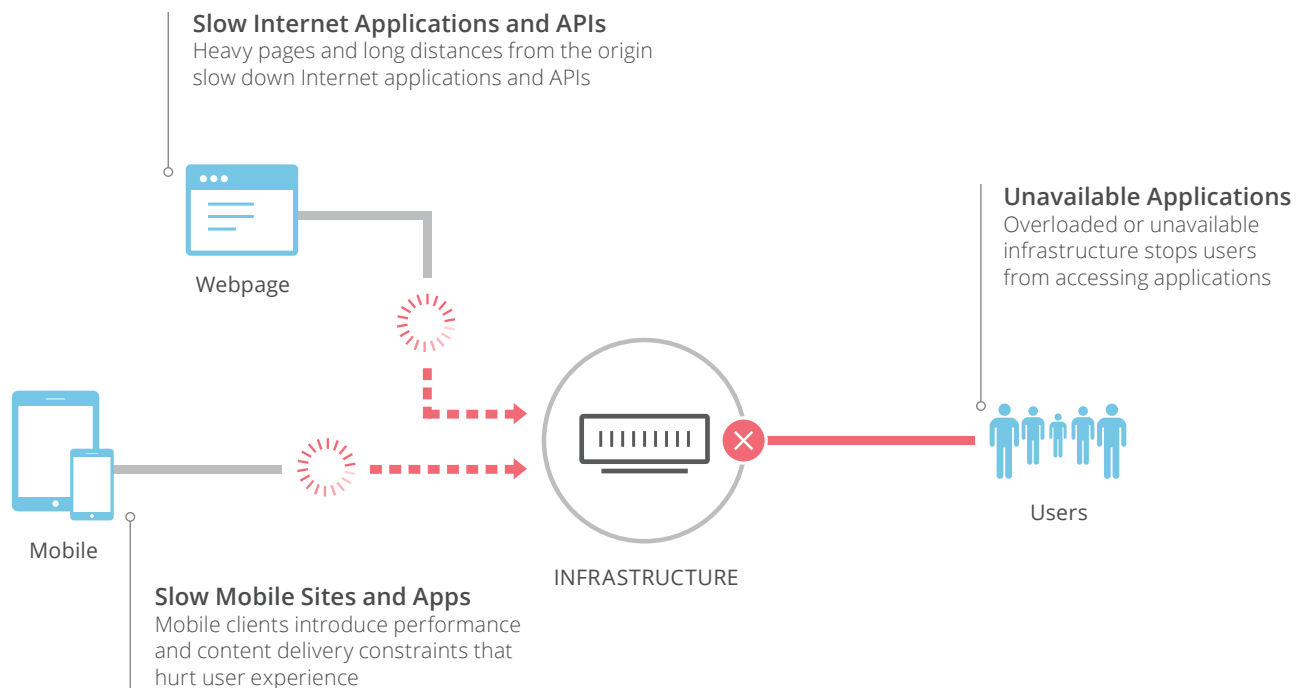
- Expectations of your customers are increasing
- Mobile usage has higher business impact than before
- Your users are expanding globally faster

Customer expectations of their applications have increased. They are demanding greater interactivity, such as more “app-like” look and feel, and more personalized content. They want “richer experiences” through more and higher quality photos, images and videos.<sup>1</sup>

As customers increase their engagement with Internet applications over mobile devices, they still expect desktop-like speeds. More engagement and transactions are being completed on mobile, surpassing desktops. According to Mary Meeker, a top analyst at venture firm KPCB, 3.1 hours were spent on average on mobile, versus 2.2 hours on desktop, per day.<sup>2</sup> The same report cited that mobile advertising in 2016 was \$37B out of the \$73B total Internet ad spend in the US. The top quartile of retailers now capture half of their sales on mobile. While all countries continue to see growth in mobile. Some countries, such as Japan and UK, sell more through mobile than desktop.<sup>3</sup>

Customers are both raising expectations and becoming more globally distributed. Shopping customers, for example, come from a wider range of countries beyond the site’s originating country. According to Nielsen, “Shoppers are increasingly looking outside their country’s borders, as more than half of online respondents in the study who made an online purchase in the past six months say they bought from an overseas retailer (57%).”<sup>4</sup> This acceleration of globalized commerce applies to nearly every business segment.

These three primary trends—increased expectations from customers, the rise of mobile, and accelerated globalization of customers and audiences—have increased pressure on application performance.



## What Impacts Application Performance?

An Internet-facing application can slow down for several reasons. Understanding causes of latency can help developers and IT professionals to design better solutions that meet their performance requirements.

The speed of going from point A, the location of a visitor, to point B, the origin or source of content, depends on a few key factors:

### **Distance**

Due to “speed of light” limitations, it will take longer for requests to travel the nearly 16,000 kilometers from New York to Sydney (80 ms) than to travel the 4,000 kilometers from New York to San Francisco (21 ms).<sup>5</sup>

### **Network throughput**

The rate at which data can flow through the network to the client will impact latency. For example, a network with 1 Mbps would take longer to deliver a 200 GB payload than a network with 100 Mbps.

### **Number of “requests”**

A page with 100 requests takes longer to load than a page with only 5 requests (assuming all request sizes and connections used are the same, with no multiplexing).

### **“Payload” size**

On a 1 Mbps network, it takes longer to deliver a 10 TB payload than a 100k file.

### **The number of requests on a single TCP connection**

Sending several requests using a single open TCP connection is faster than establishing a new TCP connection for every request.

### **The number of roundtrips to establish a secure TLS session**

Establishing a “handshake” may take several round-trips, and reducing the number of round-trips (for example, supporting [TLS 1.3](#) versus [TLS 1.2](#)) can decrease latency.

### **Client-side “render time”**

Code that needs to be processed by the client, such as decompressing images or running javascript, adds to the latency experienced by the user. Even if the payload has been delivered, client-side code must be “run” and can add time, depending on the processing power and memory available to the client.

Any one of these elements can contribute to the latency of an application. Shaving milliseconds off of any one of these potential factors can speed-up delivery of Internet applications and APIs.

## Heavier Pages and Longer Distances Slow Down Sites

Because of the demand for “richer” user experiences, Internet applications become “heavier.”

The average page size continues to increase, whether it is because of more complex CSS, larger media assets, or more JavaScript snippets. Wired, a science and technology magazine and website, reported that the average size of websites is now 2.3MB, which is larger than a famous computer game called Doom was when it was first released. The higher page size (the “payload”) increases the amount of data that must travel over the networks to be rendered by the browser.

When applications use APIs to fetch dynamic content to support personalization and interactivity, the number of connections or trips to the origin increase. For example, an interaction which fetches profile information, or processes inputs from the user around purchase preferences, increases the personalization and interactivity. Each additional trip adds latency because of the added time to make connections and the greater distances to go to the origin for dynamic content.

As users become more globally dispersed, content must travel longer distances, sometimes to countries with less reliable or lower throughput networks, which introduces more latency. A site, for example, with an origin in San Francisco may have as many visitors who are local as from halfway around the world. This increases the likelihood of global visitors experiencing a slower application.

## Mobile Experience Even More Challenging

As customers engage more on mobile, the challenges of heavier pages and longer distances to the origin servers worsen. The limited memory, CPU and power of these devices slow down the decompression or rendering of content delivered to a mobile browser, such as images or JavaScript.<sup>7</sup>

Mobile networks are often spottier than wired connections. Sometimes mobile network operators throttle connection speeds,<sup>8</sup> or in many countries, the network hasn't been upgraded to higher data rates. This means that, as a more global base of clients accesses Internet applications through mobile devices, the range of performance widens further. For example, Iraq has the slowest mobile download speed at 3.03 Mbps (Megabits per second) while Norway has the fastest at 62.59 Mbps.<sup>9</sup>

Mobile apps typically increase the number of API calls to the origin needed. This reduces the effectiveness of static caching, requiring requests to travel longer distances to the origin, while also exposing the application to malicious traffic that could overload servers and hurt performance.

Mobile continues to grow in importance to businesses wanting to increase customer engagement, conversions, and transactions. Yet at the same time, mobile devices introduce additional constraints which hurt user experience.

## Overloaded or Unavailable Applications

The other primary performance problem occurs when overloaded infrastructure degrades performance so badly that eventually the applications stop responding.

Applications can unexpectedly go offline for reasons unrelated to the network. These are often hard-to-troubleshoot outages that can involve developers, database administrators, and other IT staff to triage and resolve.

Even if the application has no issues with the code itself, traffic, both good and bad, can overload the capacity of a specific origin server or data center, making them unavailable or less performant. This problem is especially harmful when the traffic is good, such as during a holiday sale or unexpected media event, when downtime can impact revenue.

The way many companies try to mitigate these risks is through manual, error prone disaster recovery and in-house load-balancing. This increases the risks of application failure while adding maintenance, hardware, and operational costs. A common scenario with manual or in-house load-balancing is that the recovery or re-routing of traffic fails at critical times.

Sometimes the availability has nothing to do with a company's controllable assets, such as their own application or infrastructure. Instead, the networks managed by ISP or telcos are congested or error-prone, which results in timeouts or too many retries, stopping the users from accessing the Internet applications.

These common problems—**a) slow Internet applications and APIs, b) added challenges for mobile delivery, and c) overloaded or unavailable applications**—can have direct, meaningful business impact.



I estimate Argo provides us with a 5% bump in conversion.

**TOUCH OF  
MODERN**

**STEVEN OU**  
CTO

## Business Impact of Slow Applications

There are two key areas where businesses can be impacted by poorly performing applications.

- Long-term, strategic impact, such as a degraded brand and reduced lifetime customer value
- Near-term financial impact, such as lower revenue and higher operational costs.

Brands degrade when customers have a worse experience than they originally expected. One of the most common reasons for dissatisfaction with an Internet application is slow performance. According to an IDG Enterprise report, “A poor mobile website experience affects the brand reputation and makes mobile device users less likely to engage with a company/brand.”<sup>10</sup>

Once a customer has decided to stop returning to a site because of lagging load times, not only has the brand degraded through bad word-of-mouth, but lost future transactions can reduce the lifetime value from that customer.

More immediate impacts can include reduced conversions, abandoned purchases, reduced checkout size, or decreased return visits—all which can reduce revenue. In a study by Google and Bing, a mere two-second delay led to more than a 4% loss in satisfaction and, most importantly, a 4.3% loss in revenue per visitor.<sup>11</sup> Because performance problems often result from large assets and trips to the origin, operational costs from bandwidth and compute could potentially be reduced by addressing those challenges.

Given the meaningful business impacts from poor performance, what is the best approach to meeting both business goals while addressing the growing technical challenges and pressures?

“  
If we didn’t have this speed, then more people  
would drop and our bounce rate would be higher.”



RENATO MENEGASSO  
Head of Technology

## The Cloudflare Performance Advantage

To address the growing challenges and importance of performance, Cloudflare delivers an end-to-end solution with three key advantages for “always on” businesses.

The three critical differences are: scale, ease-of-use, and integration with security services.



### SCALE

Cloudflare’s 120+ data centers worldwide, each with the fully integrated stack of performance solutions including DNS, caching, and routing, provide a distributed network that delivers content to users wherever they are, on any device, as fast as possible.

This large global footprint brings content closer to customers by being in more locations. The large number of data centers also brings the Cloudflare network closer to more customer origins. This allows for more routing options that Cloudflare can optimize based on the large sample of Internet traffic from over 7M Internet applications and APIs.

The scale also makes the Cloudflare network more resilient against bad traffic. The network capacity and ability to handle the largest DDoS attacks because of Anycast ensures websites don't slow down or go offline when attacked.

## EASE OF USE

Cloudflare's UI and API enable companies to easily make changes that improve performance, either through simple manual intervention or defining programmatic policies. Technical performance-related solutions, which traditionally involved development time, additional hardware, or ongoing IT maintenance, can now be deployed and fine-tuned quickly with reduced risk of error and without reliance on a specialist or external professional services engagement.

As a result, customers without existing performance solutions realize benefits quickly because of the fast onboarding and the easy "point-and-click" deployment of acceleration features, such as smart routing and image compression. The ability to roll-out innovations, like HTTP/2 or 0 round-trip TLS 1.3 resumption, enables companies to keep pace with evolving Internet standards.

Flexible and intuitive custom caching rules allow companies to apply performance innovations to complex applications quickly with fewer errors, and in compliance with the business. Features such as fast, tag-based purging, geo-based caching and header-based caching policies allow companies to maximize the value of caching content across the network.

Because performance and availability demands round-the-clock coverage, the ability to broaden the number of users who can apply and tune performance rules helps shrink response-time to incidents.

## INTEGRATION WITH SECURITY FEATURES

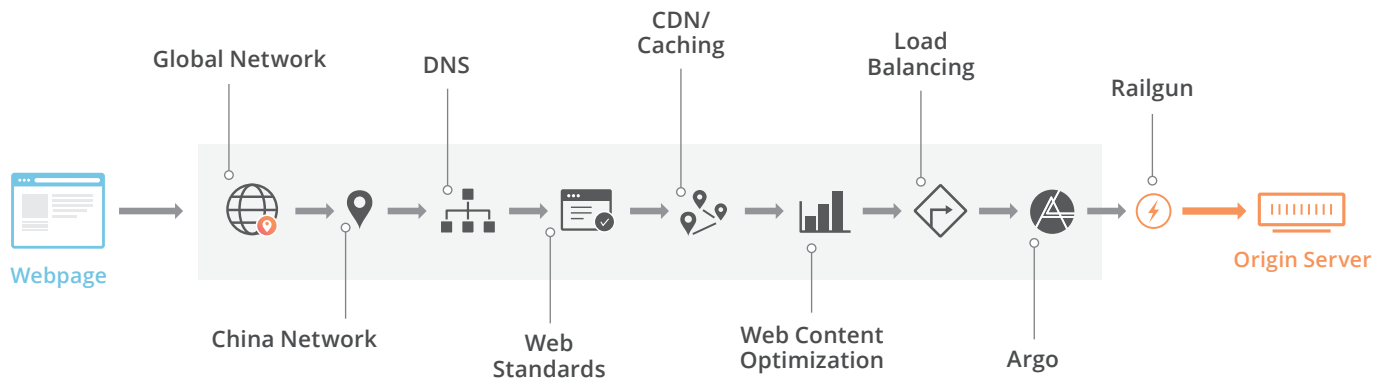
Typically, keeping sites secure has resulted in slower performance or increased complexity. Because secure connections through HTTPS introduces additional overhead for encryption/decryption, as well as additional round-trips for establishing TLS handshakes, security typically hurts performance.

By supporting the latest security protocols, like TLS 1.3, Cloudflare enables encrypted connections while reducing the handshake expense to improve the overall performance. Services like DDoS and WAF keep bad traffic, which could degrade performance, off of the origin.

## Cloudflare Web Performance Services

Each of Cloudflare's data centers provides a comprehensive stack of performance features which address the most common causes of poor performance. These features help solve many of the core problems that slow down the delivery of applications over the Internet.

From fast web address lookups to accelerated delivery to the origin server, Cloudflare speeds up traffic at key points in the life of a request.



**Cloudflare's fast DNS** shaves crucial milliseconds off requests for a DNS lookup. Delays in DNS query time could occur for every request and latency from DNS servers can slow down a site. Cloudflare's DNS, which resides in each of our globally distributed data centers, enables fast lookup times.<sup>12</sup> Cloudflare's Anycast network helps ensure faster DNS response times by answering with the DNS server closest to the visitor.

**Support for latest web standards** leverages the latest innovations to accelerate traffic that reduce the number of round trips and decrease the size of payloads. **TLS 1.3 with 0-RTT** resumption reduces the amount of time for each "handshake" to establish a secure connection. Cutting down the number of round-trips results in performance improvements.<sup>13</sup> **HTTP/2** compresses the headers, making each payload smaller, and allows a single connection to multiplex several requests in parallel, reducing the performance "expense" of total handshake round trips.

**Caching** content close to visitors reduces the distance it must travel. As a result, web sites load faster while also reducing bandwidth and infrastructure costs. Flexible caching rules to control behavior by request type, URL, and header content enable customers to tune their cache to accelerate complex applications.

**Web Content Optimization** reduces the number and the size of the assets that must travel over the Internet by supporting new browser capabilities such as *WebP* image format and pre-fetch, by compressing images, and by minifying css and javascript scripts.

**Load Balancing** with geo-steering ensures that customers are routed to the geographically nearest origin server to reduce the distance content must travel.



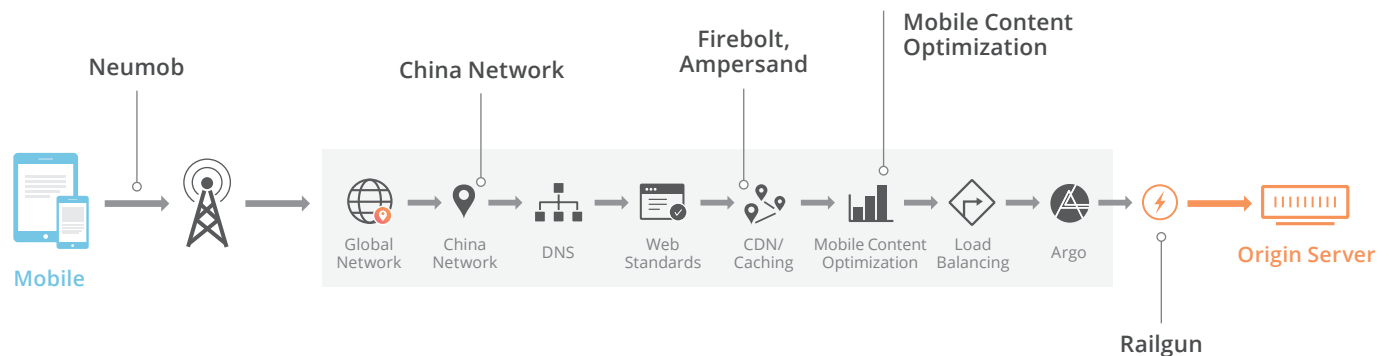
**Argo** smart routing finds the fastest, least congested path, from the visitor to the origin and accelerates how traffic is sent within the Cloudflare network. By sampling round trip times along the Cloudflare network, Argo intelligently routes traffic to ensure optimal delivery of dynamic content from the origin. Not only are the fastest routes selected, but requests are sent through a single open connection, reducing round trips and speeding up delivery of traffic.

**Railgun** accelerates dynamic or uncacheable content by compressing responses from the origin with an agent on the server.

Together, these web performance components shave off critical milliseconds for every visitor without sacrificing your customers' experience or security.

## Cloudflare Mobile Performance Services

Cloudflare's Mobile Performance Services add additional features to the above web performance capabilities to address the challenges of fast application delivery to mobile clients. Mobile delivery is challenging because mobile devices have limited CPU, memory, power, and network speeds.



**Cloudflare's China Network** allows companies to better serve their customers in China, where nearly 1.5 billion mobile users reside, 75% of whom access the Internet through their mobile phone.<sup>14</sup>

**[BETA] Neumob** accelerates mobile performance for native mobile apps while reducing errors and data loss by multiplexing requests over a custom protocol from the device to Cloudflare's data centers.

**Ampersand** gives mobile sites "app-like" responsiveness by supporting Google's Accelerated Mobile Pages (AMP). Publishers can leverage AMP to speed up delivery of mobile ads with **Firebolt**.

Cloudflare's **Mobile Content Optimization** addresses the rendering and network limitations of most mobile devices.

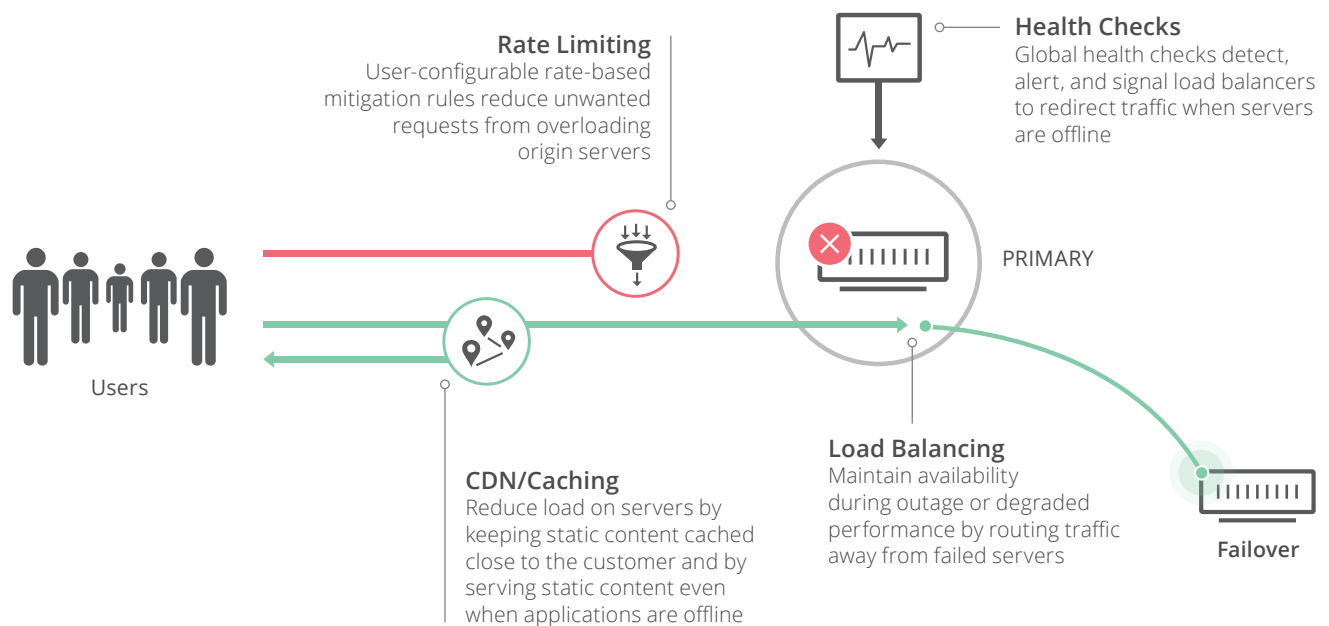
**Mirage** enhances Cloudflare’s web content optimization for mobile users by “virtualizing” images—sending them initially as a lower-resolution—and consolidating them into a single request. Reducing size and connection requirements addresses the limited power, memory and CPU for mobile phones compared to desktops. Because of the throughput constraints for most mobile networks, **HTTP/2**, by keeping a single, open connection and multiplexing several requests along that connection, can dramatically improve mobile delivery from 3-15 times faster.<sup>15</sup>

**Argo** accelerates mobile API calls to the origin by optimizing routes and keeping connections open along the Cloudflare network.

**Railgun** speeds up native mobile apps by compressing the server response to APIs.

## Cloudflare Availability Services

Cloudflare improves availability of applications by protecting the origin from being overloaded and redirecting traffic around congested networks and away from failed servers to healthy origins.



**Caching** enables applications to continue to serve static content, even when their origin servers are unavailable. This allows their customers to continue to access and consume pages even when the application is offline. It also reduces risk of downtime by keeping requests for static content off the origin.

**Load Balancer** uses **global health checks** to identify servers that are down, and route traffic to another available server or data center to ensure application availability. Customers can define their own pools of servers as well as the thresholds and locations of health checks to keep applications online and available. Automatic failover minimizes downtime, and alerts on servers that are down help speed up problem discovery and failure resolution.

Integrated security features, like **DDoS protection and WAF**, keep attack traffic off the network and prevent malicious actors from bringing a site or application down.

## Summary

Cloudflare accelerates Internet applications and APIs across its large global network to both mobile and desktop users. It is easy to use, ensures application availability and delivers fully integrated security across its entire network of datacenters. Cloudflare provides predictable cost control for companies with flat rate enterprise pricing.

## References

- <sup>1</sup> Forbes, "[Top Web Design Trends To Watch In 2017](#)", January 25, 2017, Tomas Laurinavicius
- <sup>2</sup> Mary Meeker's Internet Trends Report, 2017 <http://www.kpcb.com/internet-trends>
- <sup>3</sup> [Criteo Mobile Commerce Report, 2016](#)
- <sup>4</sup> "Global Connected Commerce Is e-tail therapy the new retail therapy?", Nielsen, January 2016
- <sup>5</sup> Time is based on light traveling along fiber
- <sup>6</sup> "[The Average Webpage Is Now the Size of the Original Doom](#)", Klint Finley, 4.23.16, Wire  
*"Today's average webpage, meanwhile, requires users to download about 2.3MB worth of data, according to HTTP Archive, a site that tracks website performance and the technologies they use."*
- <sup>7</sup> <http://sealedabstract.com/rants/why-mobile-web-apps-are-slow/>
- <sup>8</sup> MobiForge, "You've been throttled, but don't stop browsing," December 15, 2017, Ruadhan O'Donoghue, <https://mobiforge.com/news-comment/youve-been-throttled-dont-stop-browsing>
- <sup>9</sup> "Speedtest Ranks Internet Access Speed In More Than 100 Countries," Forbes, August 14, 2017, Kevin Murnane, <https://www.forbes.com/sites/kevinmurnane/2017/08/14/speedtest-ranks-internet-access-speed-in-more-than-100-countries/#46edc1d065b5>
- <sup>10</sup> IDG Enterprise/IGS Mobility Survey 2014, October 29, 2014, <https://www.idgenterprise.com/resource/research/2014-idg-enterpriseigs-mobility-survey/>
- <sup>11</sup> Schurman, Eric, Bing and Brutlag, Jake, Google Search, "Performance Related Changes and their User Impact" 2009 Velocity Conference
- <sup>12</sup> DNSPerf speed test, November 27, 2017, trailing 30 days, shows Cloudflare 15.98 ms and large DNS cloud providers averaging 61.21 ms, 81.52 ms, and 77.58 ms
- <sup>13</sup> <https://blog.cloudflare.com/introducing-0-rtt/>
- <sup>14</sup> <https://www.statista.com/statistics/278204/china-mobile-users-by-month/>
- <sup>15</sup> <https://medium.com/apps-and-networking/http-2-makes-media-loading-3-15-times-faster-on-mobile-a455c3e68135>



1 888 99 FLARE | [enterprise@cloudflare.com](mailto:enterprise@cloudflare.com) | [www.cloudflare.com](http://www.cloudflare.com)

---

© 2017 Cloudflare Inc. All rights reserved.  
The Cloudflare logo is a trademark of Cloudflare. All other company and product names may be trademarks of the respective companies with which they are associated.