

The Bandwidth Madness



Dipl. Ing. (FH) Klaus Rock

Bonhoefferstrasse 37

D - 73432 Aalen

Germany

E-Mail: : k.rock@rock-technologies.com

1 WHY BANDWIDTH DOESN'T MATTER

Why is it whenever you hear people talking about speed on the internet, they use a single metric? Whether they're discussing the connection in the datacenter, their residential DSL, or the wireless connection via their mobile device, everyone references the bandwidth of their connection when talking about speed. "Oh I just got a 20Mb/s connection at home, it's blazing fast!". That's all well and good, and 20 Mb/s is indeed a lot of throughput for a residential connection. Unfortunately for Joe Average, about 98% of the population wouldn't know what the heck to DO with 20 Mb/s of download speed, and even worse than that...they would likely see absolutely **zero increase** in performance while doing the one thing most people use their connection for the most, browsing the web.

No one seems to ever bother mentioning the true culprit for slow (or fast) web browsing performance: **latency**, measured in **RTT** (Round Trip Time). This is the measure of time it takes for your system to make a request to the server and receive a response back. I.E. one complete request loop. **This is where the battle for speed when browsing the web is won and lost.** A round trip is measured in milliseconds (ms). This represents how much time it will take regardless of file size (this is important later in the discussion) to make the trip from you to the server and back. This means each connection you have to open with the server for an additional request must take at least this long. You add in the time it takes to download each file after accounting for RTT.

"Impossible!" you say, "Clearly going from a 10Mb/s connection to the new, fast, fancy (expensive?) 20 Mb/s connection my provider is proud to now be offering will double my speed on the web! 20 is twice as much as 10 you dullard!" you assert? Oh how wrong you are, dear uneducated internet user. Allow me to illuminate the situation via a brief discussion of what actually occurs when you are browsing the web. We'll skip some of the fine grained details and all the DNS bits, but here's the general idea:

Whenever you make a request for a web page on the net, you send out a request to a server. That server, assuming you're an allowed user, then sends a response. Assuming you don't get redirected and are actually served a page, the server will send you a generally simple HTML page. This is a single, small file that contains the HTML code that tells your browser how to render the site. Your computer receives the file, and your browser goes to work doing exactly that, rendering the HTML.

Up to this point people tend to understand the process, at least in broad strokes. What happens next is what catches people I think. Now that your browser is rendering the HTML, it is not done loading the page or making requests to the server, not by a long shot. You still haven't downloaded any of the images or scripts. The references to all of that are contained in the HTML. So as your browser renders the HTML for the given site, it will begin sending requests out to the server asking for those bits of content. It makes a new request for each and every image on the page, as well as any other file it needs (script files, CSS files, included HTML files, etc.)

Here are the two main points that need to be understood when discussing Bandwidth vs. RTT in regards to page load times:

1. The average web page has over 50 objects that will need to be downloaded (reference: <http://www.websiteoptimization.com/speed/tweak/average-web-page/>) to complete page rendering of a single page.
2. Browsers cannot (generally speaking) request all 50 objects at once. They will request between 2-6 (again, generally speaking) objects at a time, depending on browser configuration.

This means that to receive the objects necessary for an average web page you will have to wait for around **25 Round Trips** to occur, maybe even more. Assuming a reasonably low 150ms average RTT, that's a full 3.75 seconds of page loading time not counting the time to download a single file. That's just the time it takes for the network communication to happen to and from the server. Here's where the bandwidth vs. RTT discussion takes a turn decidedly in the favor of RTT.

See, the file size of most of the files necessary when browsing the web is so minute that bandwidth really isn't an issue. You're talking about downloading 30-60 tiny files (60k is an average). Even on a 2Mb/s connection which would be considered extremely slow by today's standards these files would each be downloaded in a tiny fraction of a second each. Since you can't download more than a few at a time, you couldn't even make use of a full 2 Mb/s connection, in most situations. So how do you expect going from 10Mb/s to 20Mb/s to actually increase the speed of browsing the web when you couldn't even make use of a 2Mb/s connection? The answer is: You shouldn't.

Sure, if they were downloading huge files then bandwidth would be king, but for many small files in series, it does almost nothing. You still have to open 50 new connection, each of which has a built in 150ms of latency that can't be avoided before even beginning to download the file. However, if you could lower your latency, the RTT from you to the server, from 150ms down to 50ms, suddenly you're shaving a full 2.5 seconds off of the inherent delay you're dealing with for each page load. Talk about snappier page loads...that's a huge improvement.

Now of course I realize that there are lots of things in place to make latency and RTT less of an issue. Advanced caching, pre-rendering of content where applicable so browsers don't have to wait for ALL the content to finish downloading before the page starts rendering, etc. Those are all great and they help alleviate the pain of higher latency connections, but the reality is that in today's internet using world bandwidth is very rarely a concern when simply browsing the web. **Adding more bandwidth will not, in almost all cases, increase the speed with which you can load websites.**

Bandwidth is king of course, for multi-tasking on the web. If you're the type to stream a video while downloading audio while uploading pictures while browsing the web while playing internet based games while running a fully functioning (and legal) torrent server out of your house...well then...you might want to stock up on bandwidth. But don't let yourself be fooled into thinking that paying for more bandwidth in and of itself will speed up internet browsing in general when only performing that one task.

2 MICROSOFT IS INTENTIONALLY RESTRICTING INTERNET DOWNLOAD SPEEDS

I have always noticed that download speeds from my server were not as fast as they 'should' have been from my Laptop (given my Internet speed and the fact that other PC's were fast), but never spent the time to find out why. That all changed when Comcast upgraded my Internet download speed to around 120Mbps. I wanted that speed from my laptop! It was time to find the root cause of the slow download speeds. After days of sleuthing and days of research, I finally found out why, and the result was nothing short of stunning: **Microsoft is intentionally restricting Internet download speeds!**

What determines/limits download speed? There are many factors that determine actual download speed, but if we assume that a client and server are both on fast links, the critical factor that will absolutely determine (and possibly limit) download speed is something called the 'TCP receive window' (or RWIN), which is set on a per-socket connection. Additionally, there are ways to 'scale' this receive window size (how does not matter, but it explains some terminology seen below). Windows (Vista/7/8/etc) will automatically set -- and more importantly, increase -- the size of the TCP receive window for you, as needed, to maximize throughput.

Receive Window Auto-Tuning

Microsoft calls this automatic management of the receive window size 'auto-tuning'.

The reason why most people will NEVER notice this problem

The issue only happens on fast Internet connections: In Windows "restricted" Auto-Tune mode, Windows really does not like the RWIN to go above 256KB (262144) bytes. And if you know the RTT for the server you want to download from, that defines a maximum Mbps. As long as that maximum is above your rated Internet speed, you will not even notice this problem!

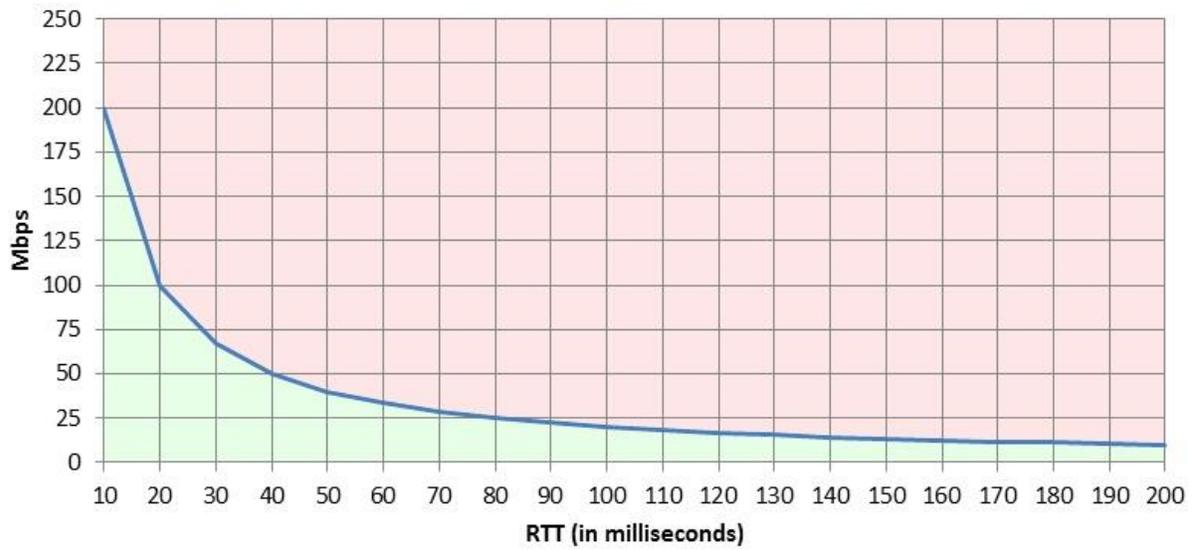
The average broadband user can NOT run into this problem: In mid 2014, the 'average' U.S. broadband speed was around 30Mbps (source: <http://www.netindex.com>). At that speed, you would have to download from a server around 70ms RTT away to even notice the problem. Most people in the US, connecting to a server in the US, will experience RTT less than that. The problem will never be seen for most people.

Fast (low RTT) servers also hide the problem: Virtually all 'big' companies (Apple, Microsoft, etc) use cloud based servers, meaning that servers are usually only around 15ms away from you. At that RTT, to even notice this problem, you would need an Internet connection faster than 133Mbps.

See on the next page a chart of $RWIN/(RTT/1000)$, when RWIN is 256K, and millisecond RTT

Take your (rated) broadband internet connection speed, and take the RTT for a server that you are downloading from. Is that intersection above the line in the red (you will see this problem), or it that intersection below the line in the green (you will not see this problem).

Microsoft Windows maximum download speed for "restricted" Auto-Tuning



As you can see as higher the Round Trip Time (RTT) as lower the available Bandwidth.

So RTT destroys Bandwidth!!