26 Jun 2023

NOTICING WHEN AN APP HAS SERVERS IN DIFFERENT REGIONS

I was talking to a friend of mine about native apps and how some apps not built using UIKit/SwiftUI just feel laggier, and it got me thinking about other latency and small performance things you notice.

It reminded me about how every time I land back in Europe or in Asia, I can immediately tell which apps have a global CDN or edge computing. Everything just immediately feels a little slower.

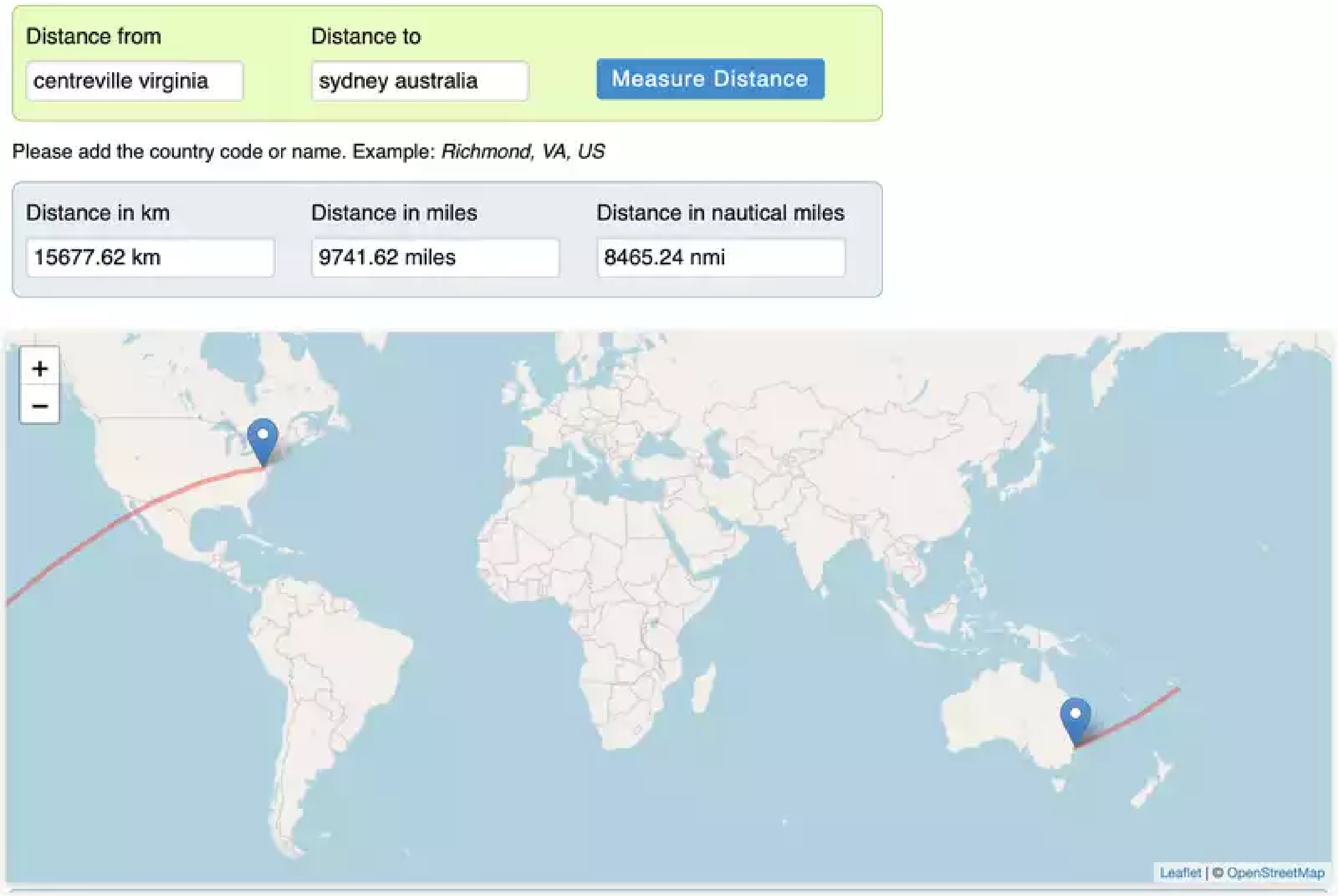
There’s some debate about when users start feeling latency, but it’s widely accepted that increased latency will impact conversions and user behavior significantly.

The figure often thrown around is that for every 100ms of latency amazon lost ~1% of sales. While I don’t know how true this is (especially given how slow the amazon app feels), it is something I know first hand. While at Pinterest we spent a ton of time improving the performance of our landing pages, as the faster it was the more users would convert and the more content they would consume.

Floored Latency

The speed of the experience you can offer your users is floored by a few variables, but most notably the physical distance from the origin server to where the user is actually sitting.

Us-east-1 (appropriately located right by “Centreville, Virginia”) is one of the main data centers run by AWS. If you’re a startup (or even quite a few mature companies), you are likely to deploy your application here. If you manage a stateful service, or your architecture doesn’t support distributed compute, you are likely to *only* deploy here. If you have users in Sydney, Australia, any network request you make will need to travel 15,677km to make it there, as the crow flies.



Distance from
centreville virginia

Distance to
sydney australia

Please add the country code or name. Example: *Richmond, VA, US*

Distance in km	Distance in miles	Distance in nautical miles
15677.62 km	9741.62 miles	8465.24 nmi

Distance from Centreville, Virginia to Sydney, Australia

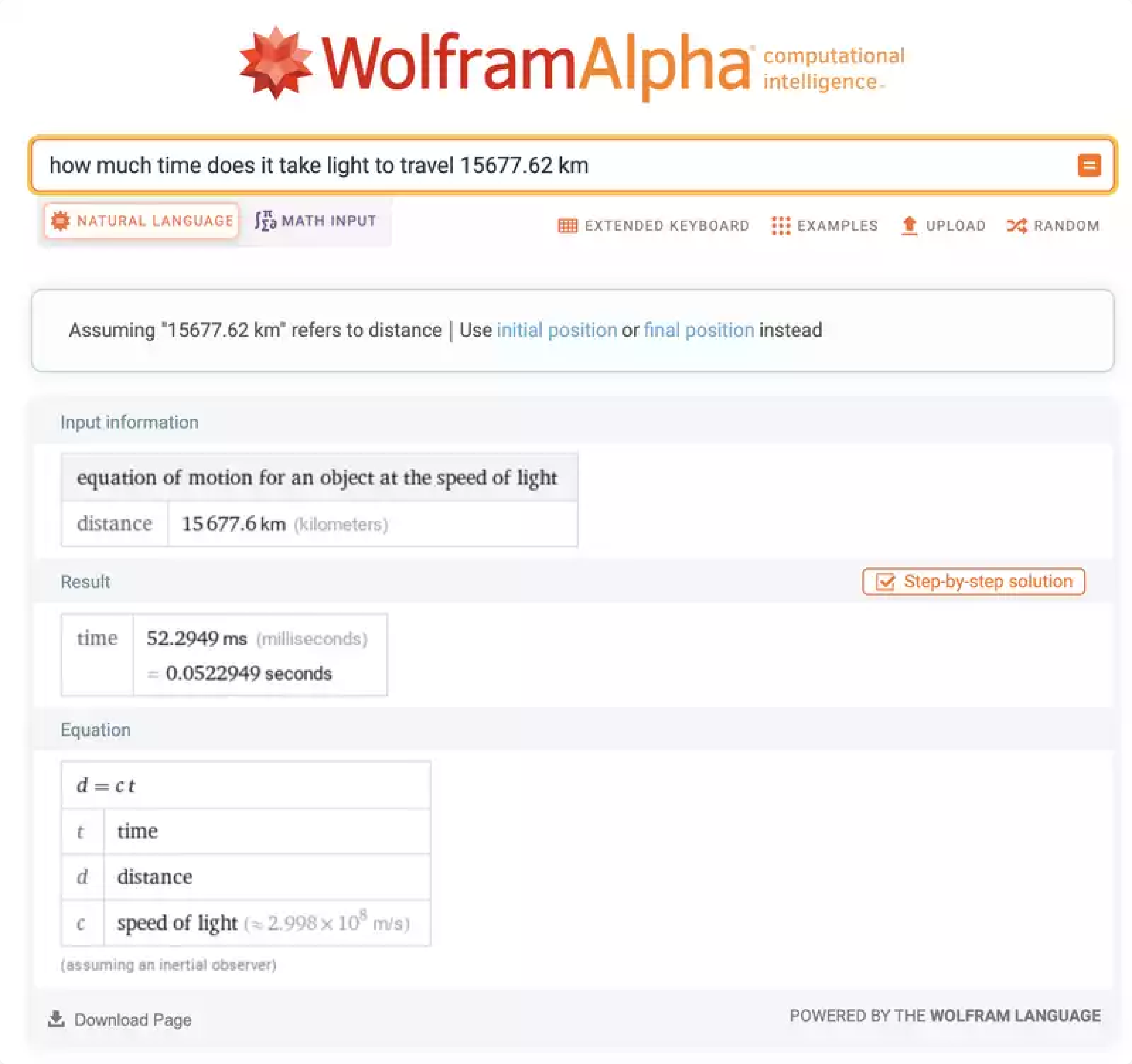
If you’re traveling at:

1) The speed of light

2) as the crow flies

3) with no overhead

then that means that you are *floored* at 104ms of latency for your request.



how much time does it take light to travel 15677.62 km

Assuming "15677.62 km" refers to distance | Use initial position or final position instead

Input information

equation of motion for an object at the speed of light	
distance	15677.6 km (kilometers)

Result

time	52.2949 ms (milliseconds) = 0.0522949 seconds

Equation

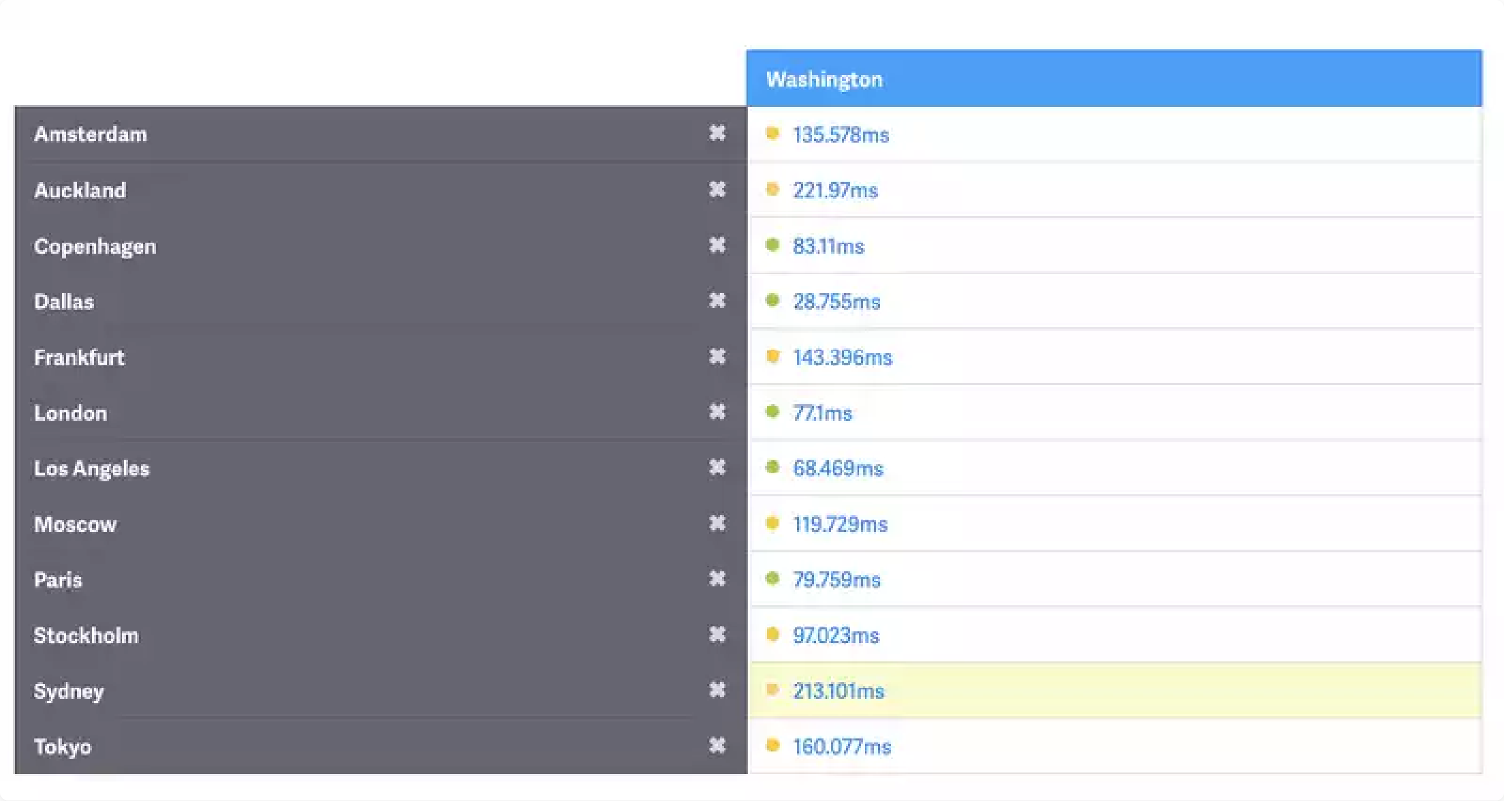
$d = c\,t$	
t	time
d	distance
c	speed of light ($\approx 2.998 \times 10^8$ m/s)

(assuming an inertial observer)

Time in takes for the speed of light to travel from us-east-1 to Sydney

And this is assuming no interference, other traffic, or time spent handling the request.

In reality, the ping you'll experience will be worse, at around 215ms (which is a pretty amazing feat in and of itself - all those factors above only double the time it takes to get from Sydney to the eastern US).



	Washington
Amsterdam ✖	● 135.578ms
Auckland ✖	● 221.97ms
Copenhagen ✖	● 83.11ms
Dallas ✖	● 28.755ms
Frankfurt ✖	● 143.396ms
London ✖	● 77.1ms
Los Angeles ✖	● 68.469ms
Moscow ✖	● 119.729ms
Paris ✖	● 79.759ms
Stockholm ✖	● 97.023ms
Sydney ✖	● 213.101ms
Tokyo ✖	● 160.077ms

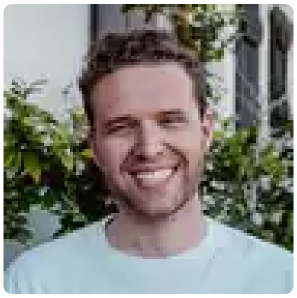
Ping to various cities around the globe

Realized latency

Having spent so much time trying to optimize web pages and API responses for performance, I've gotten a pretty good internal model for latency. I can't quite tell the difference between a us-east-1 server when I'm in New York versus San Francisco, but I can definitely tell if you've got an instance deployed in `eu-central-1` or not when I'm in Italy, or `ap-east-1` when I'm in Sydney.

Using a global CDN can help get your assets to your users quicker, and most companies by this point are using something like Cloudflare or Vercel, but many still only serve static or cached content this way. Very frequently the origin server will still be a centralized monolith deployed in only one location, or there will only be a single database cluster.

As soon as you land back in the United States and turn of Airplane mode on your phone everything just starts feeling… snappier? A little more fluid? As much as T-Mobile and Verizon would like to take credit for that I don't think theres much more to it than the physical location of the servers and where you are at that moment.



JonLuca at 12:26

To get notified when I publish a new essay, please subscribe here.