

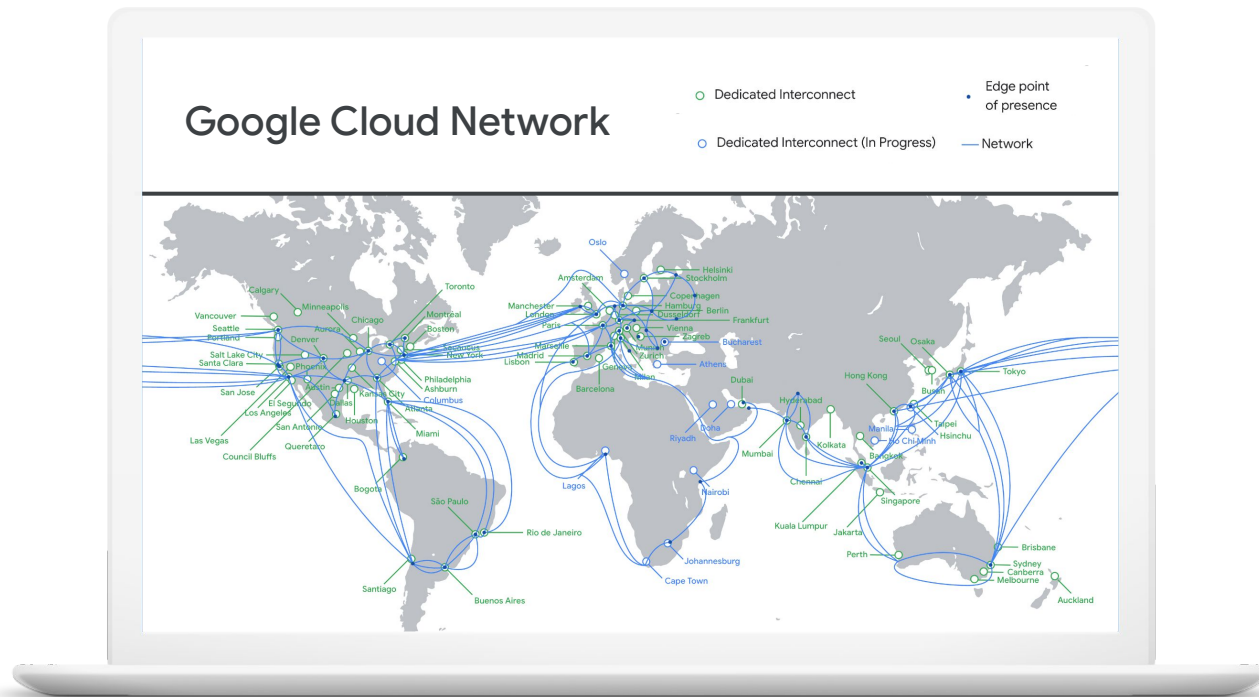
# | 10 lessons in running the largest, bestest WAN

Or, how I learned to stop worrying and love the chaos.

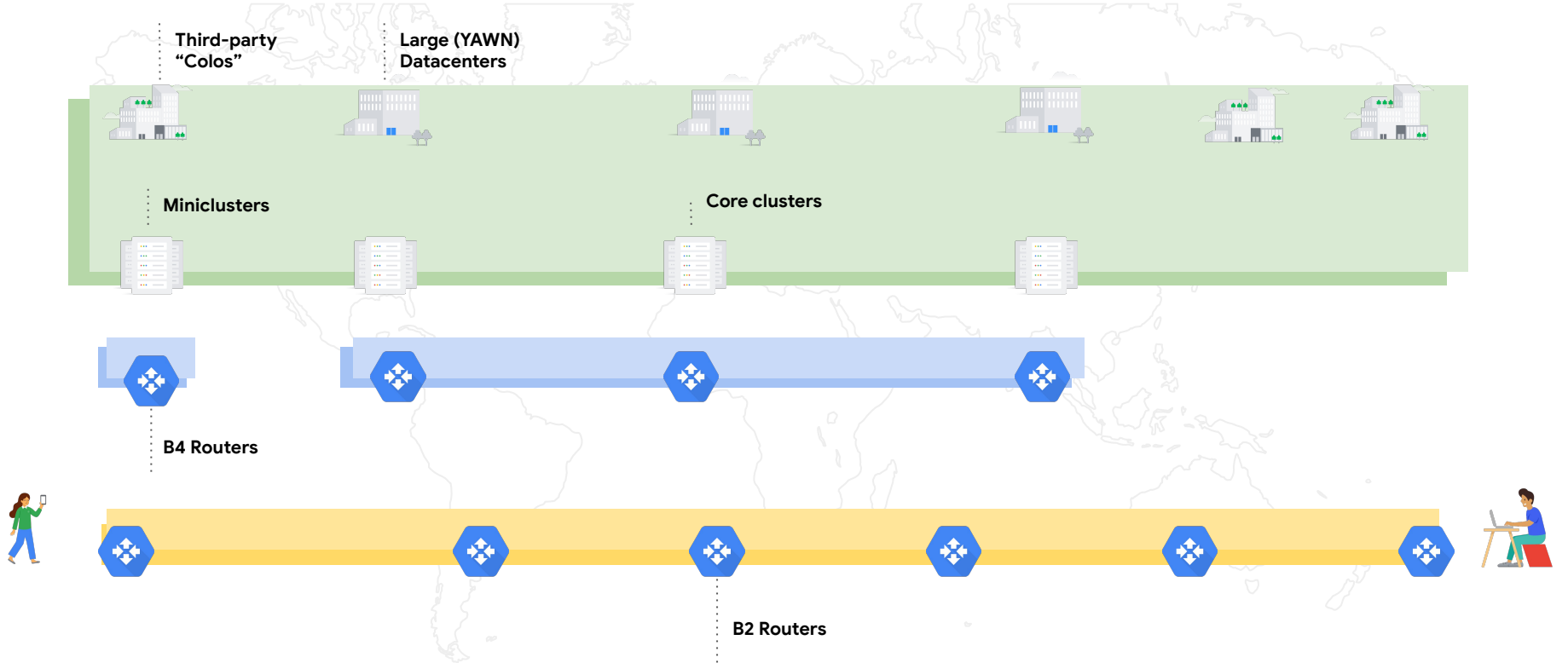
Ashok Narayanan, [ashokn@google.com](mailto:ashokn@google.com), Rob Shakir, [robjs@google.com](mailto:robjs@google.com)

## B2 - Google's global WAN

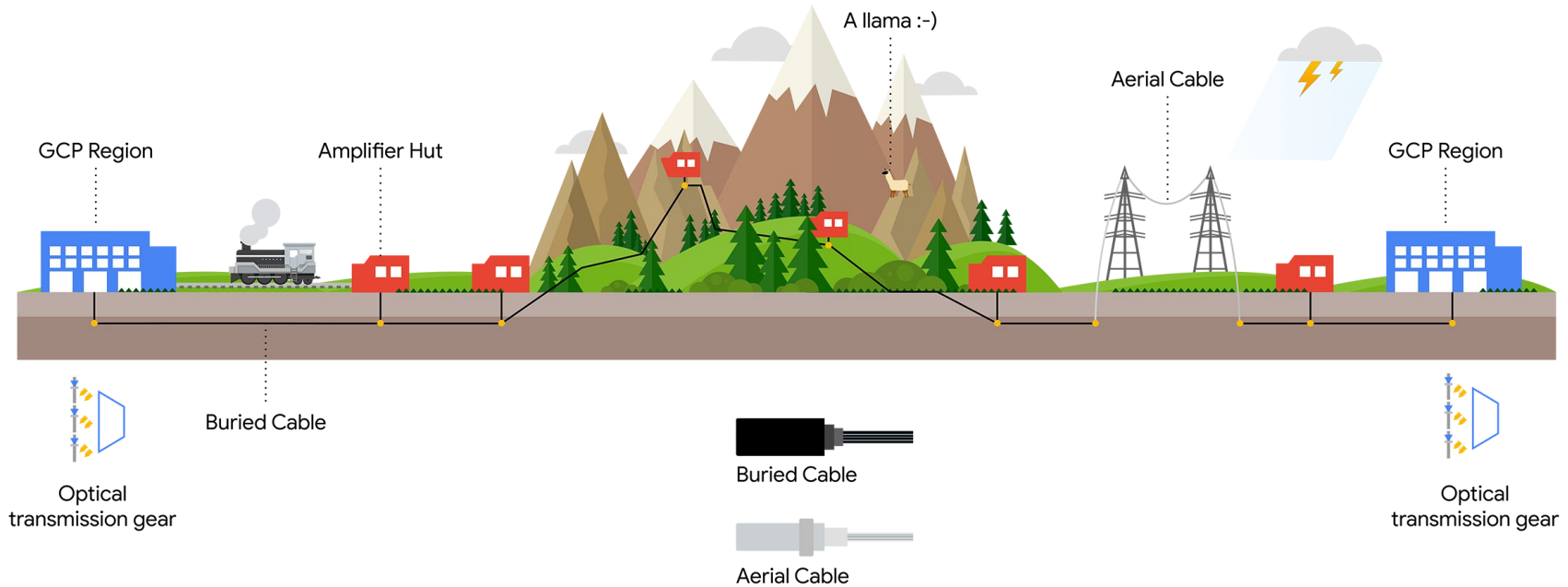
- **35** regions
- **106** zones
- **173** network edge locations
- **22** subsea cables
- **200+** countries



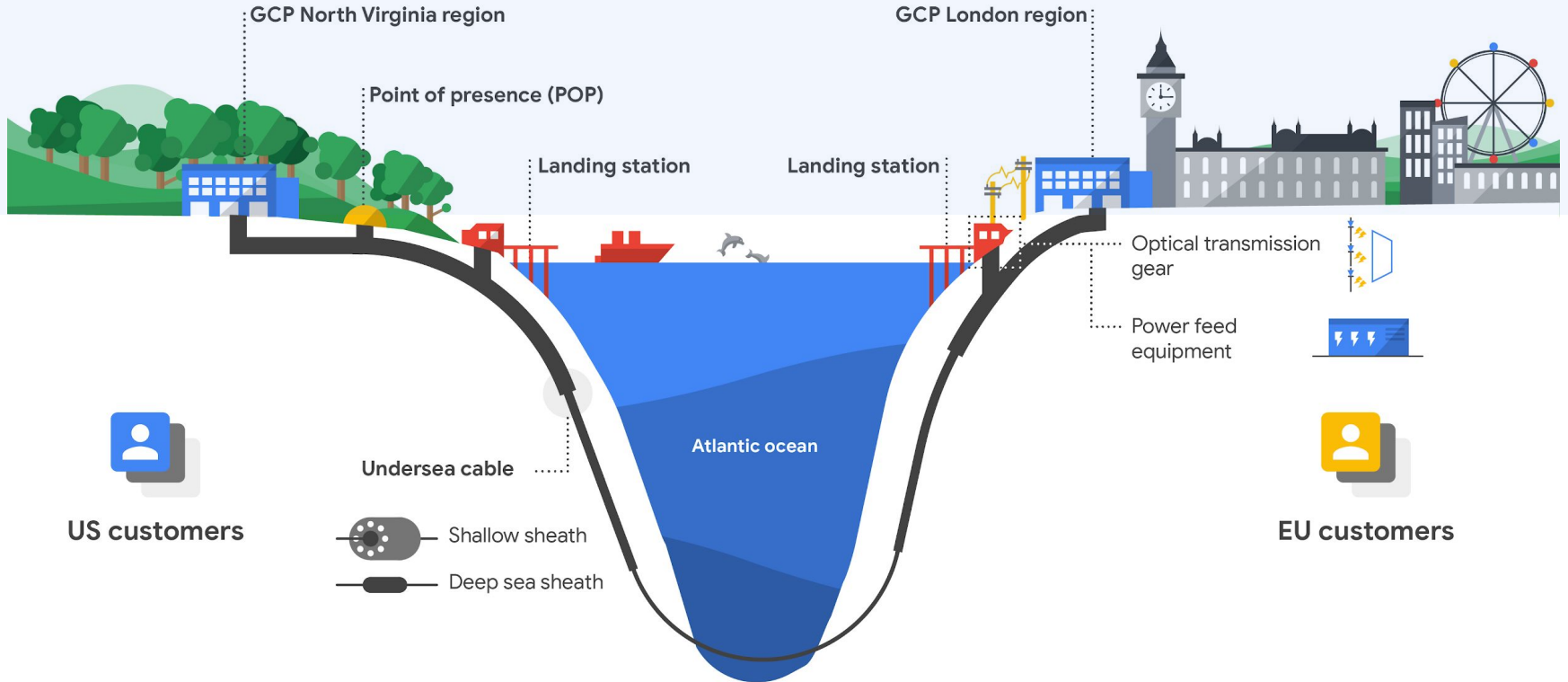
# What Googlers see...



# ...under the ground...

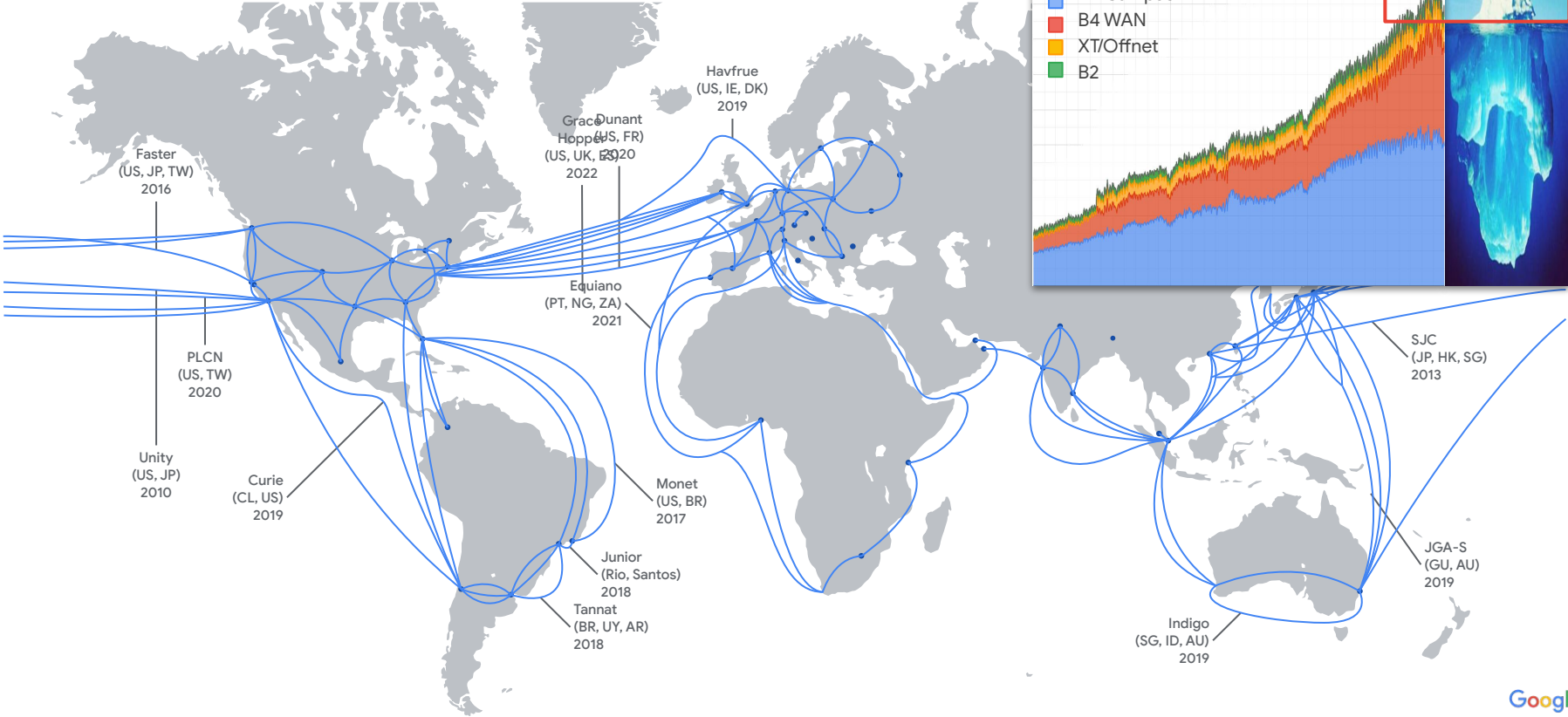


# ...and under the ocean.



# 2x Internet Scale!

Externally visible  
12% of internet



01

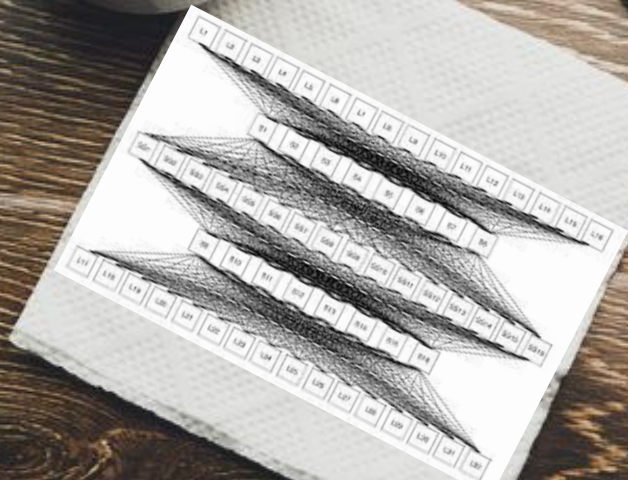
Necessity is the mother  
of invention

*... so don't let the impossible stand in the  
way of the necessary*



This Googler drew a picture of a data center network on a napkin in 2003 as a blueprint for a 10,000 port data center network he wanted to build.

All networking vendors came back with blank bids, asking us to check if we had an “extra zero” in the port count.

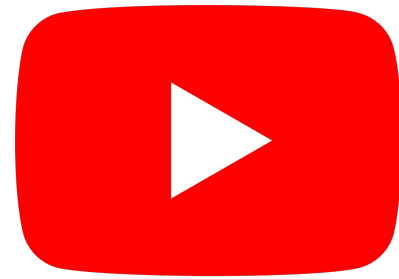




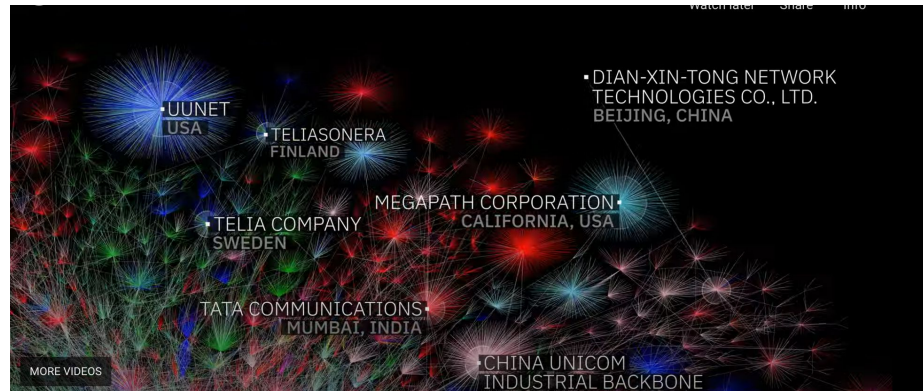
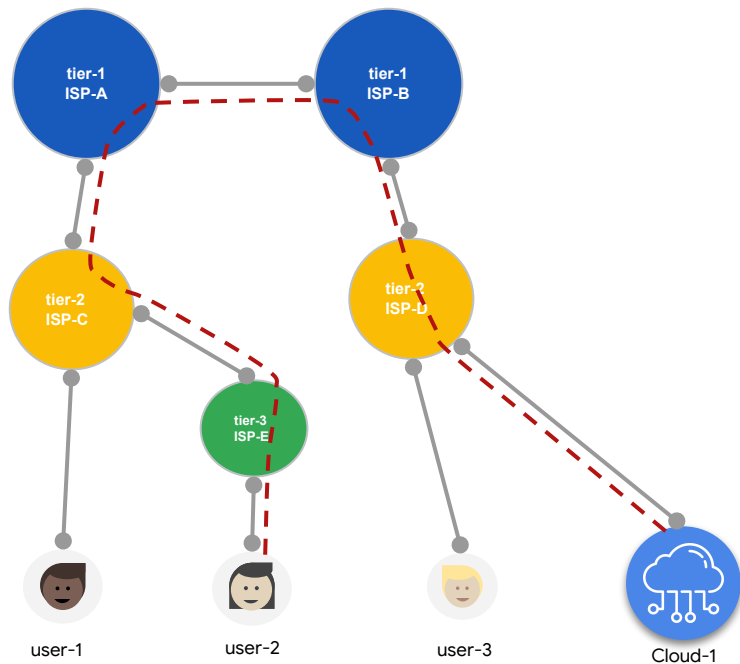
# | A simple goal

Deliver unlimited, free, high quality video to anyone, anytime, anywhere in the world\*

*\* and try not to go out of business while doing it*



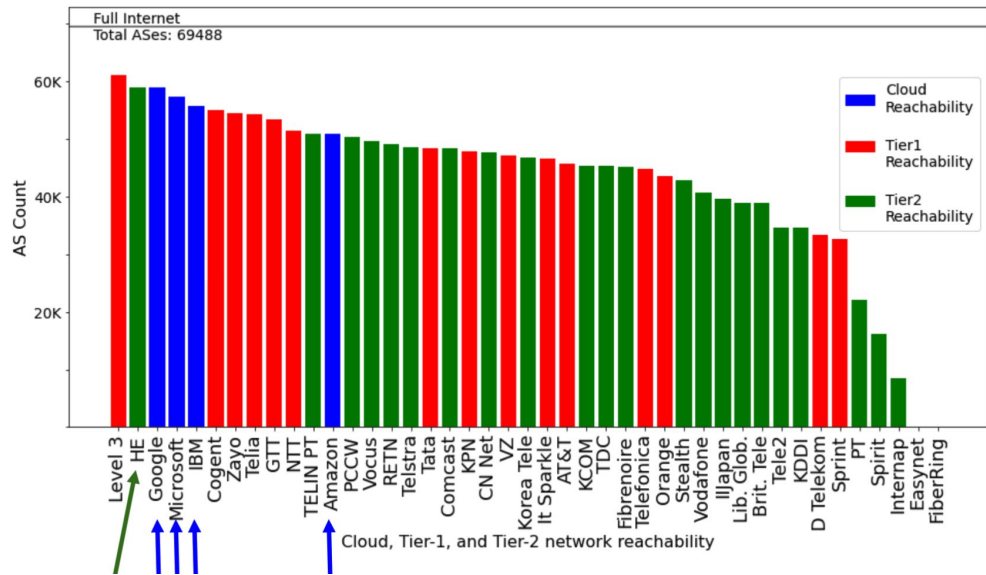
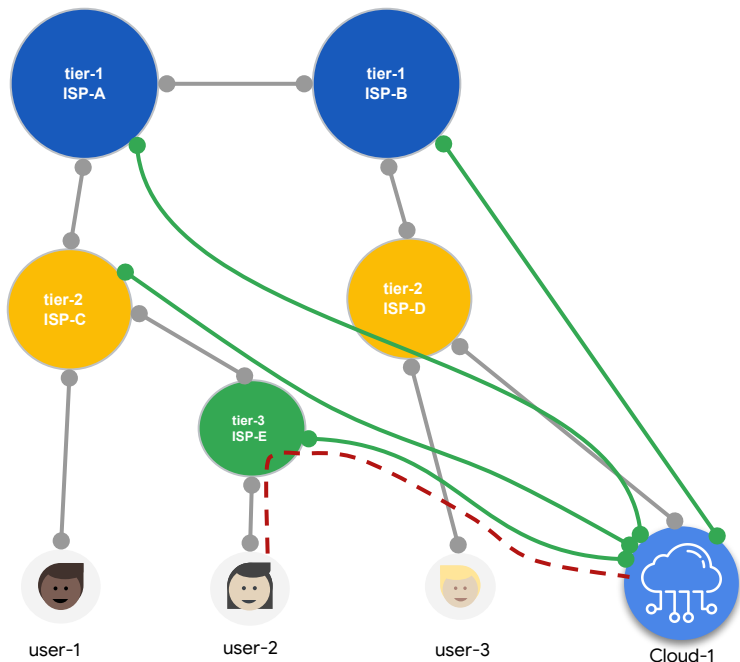
# | ... with a lot of middlemen to get through



Internet: 2007

Ref: <https://www.opte.org/the-internet>

# Flattening the Internet



## Internet: 2021

# | Gold plated peering...

- This is our (circa 2013) peering edge router, Juniper MX960
- Cost for 100 Tb/s of ports in 2013: **\$60 million**
- ... plus, you have to build a backbone network with that capacity.
- We needed to figure out a way to **reduce this cost by 10x**

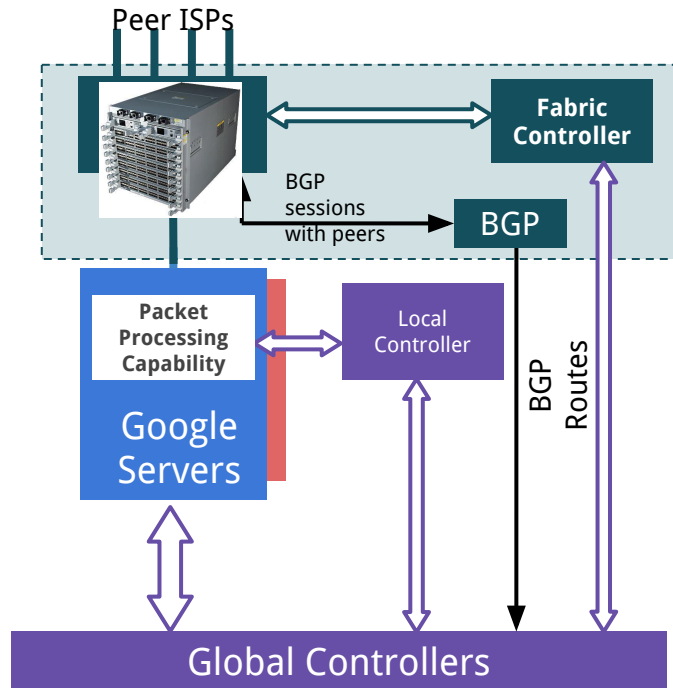


## Juniper MX960



\$60 million for 100Tb/s of peering capacity

## Espresso [Sigcomm '17]

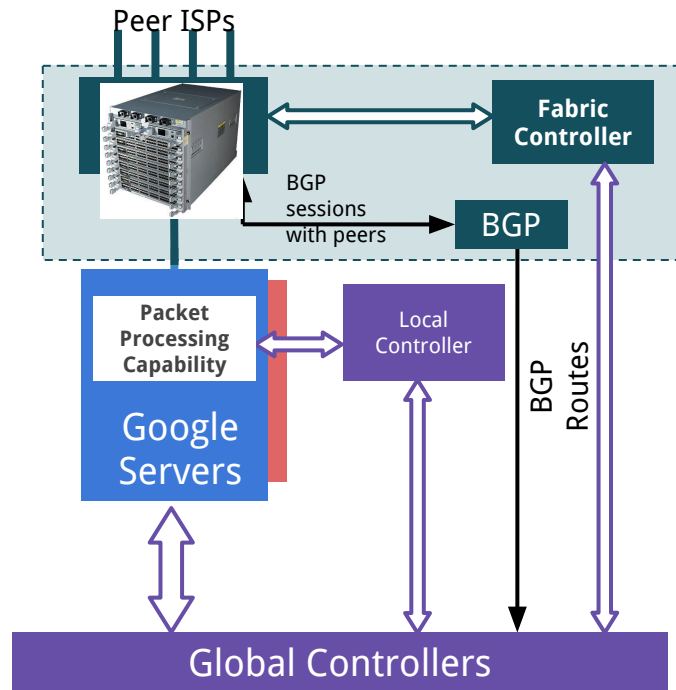


\$7 million per 100Tb/s of peering capacity

# In the end, it's all just software

- Cheaper hardware, with some packet forwarding features
- Disaggregated design with device controlled by Google software
- Peers directly connect to our SW

**Cost for 100 Tb/s of ports: \$7 million**



02

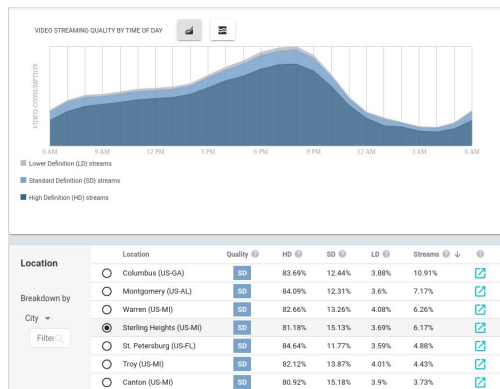
Sometimes, it's not just  
the tech that needs  
changing

# | Our (truly) Global Edge



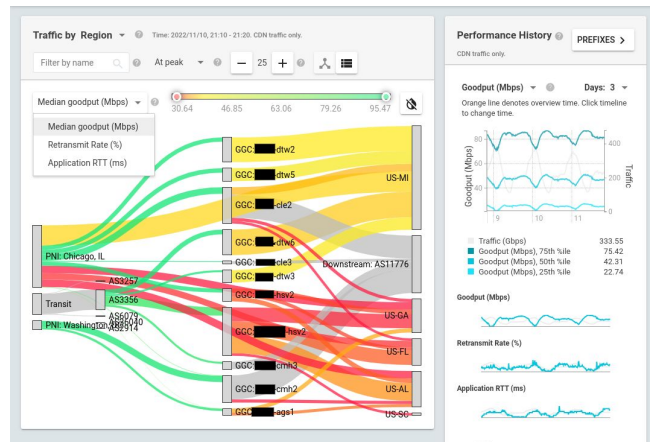
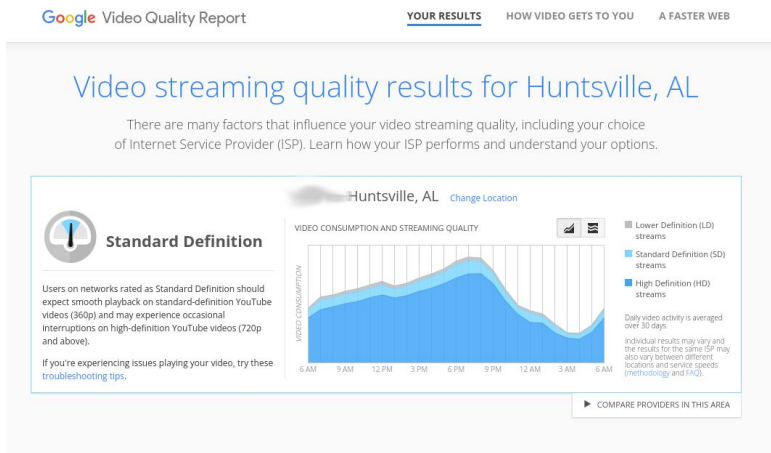


# Public ... enemy?



Use the public as your ally, by publishing a Video Quality Report ...

... but support your ISP partners, by giving them the tools they need to identify and fix performance issues



# User View

Google Video Quality Report

**YOUR RESULTS**

HOW VIDEO GETS TO YOU

A FASTER WEB

## Video streaming quality results for Huntsville, AL

There are many factors that influence your video streaming quality, including your choice of Internet Service Provider (ISP). Learn how your ISP performs and understand your options.



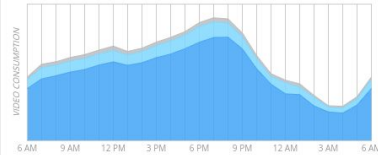
### Standard Definition

Users on networks rated as Standard Definition should expect smooth playback on standard-definition YouTube videos (360p) and may experience occasional interruptions on high-definition YouTube videos (720p and above).

If you're experiencing issues playing your video, try these troubleshooting tips.

Huntsville, AL [Change Location](#)

#### VIDEO CONSUMPTION AND STREAMING QUALITY



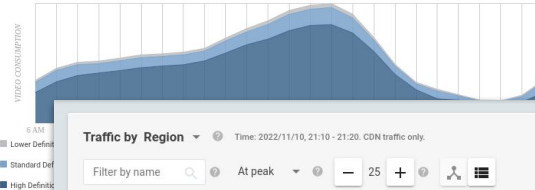
- Lower Definition (LD) streams
- Standard Definition (SD) streams
- High Definition (HD) streams

Daily video activity is averaged over 30 days. Individual results may vary and the results for the same ISP may also vary between different locations and service speeds (methodology and FAQ).

[COMPARE PROVIDERS IN THIS AREA](#)

# ISP View

#### VIDEO STREAMING QUALITY BY TIME OF DAY



Traffic by Region Time: 2022/11/10, 21:10 - 21:20. CDN traffic only.

Filter by name



At peak



25



Median goodput (Mbps)



Median goodput (Mbps)

Retransmit Rate (%)

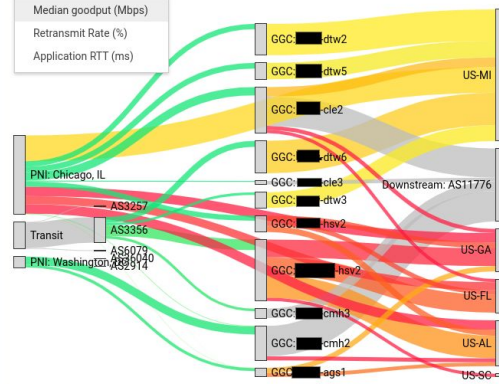
Application RTT (ms)

Location

Breakdown

City

Filter



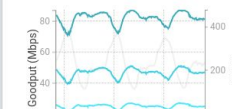
Performance History

PREFIXES

CDN traffic only.

Goodput (Mbps) Days: 3

Orange line denotes overview time. Click timeline to change time.



Traffic (Gbps) 333.55  
Goodput (Mbps), 75th %ile 75.42  
Goodput (Mbps), 50th %ile 42.31  
Goodput (Mbps), 25th %ile 22.74

Goodput (Mbps)

Retransmit Rate (%)

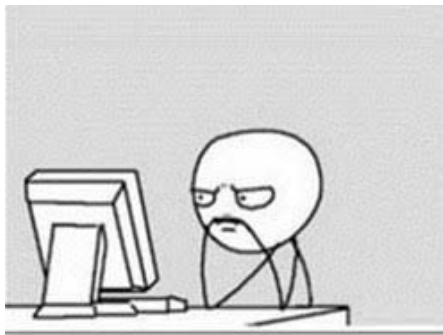
Application RTT (ms)

# 03

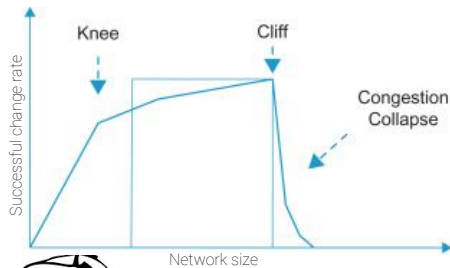
Any manual process  
done on a large network  
is guaranteed to fail

*... and so are automated processes, for  
that matter*

# Hi, my error rate is ... slim?



Human error rates range from 1% for routine tasks, to >10% for complicated non-routine tasks [Smith DJ et al]



Change rate compared to network size, if actuated by humans

Adding more humans to build a bigger network only works to a point, after which errors and delays choke the system.

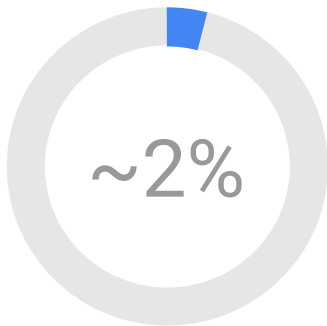
**CHALLENGE ACCEPTED**



As the scale of the network grows, and the rate of growth accelerates, **automation is a must.**

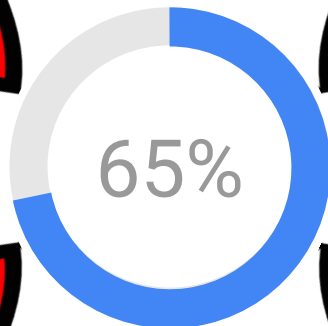
# Correlated failures

What would cause parallel devices to “fail” at the same time?



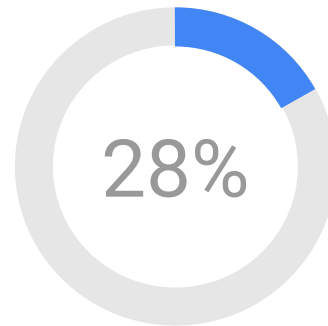
**Meteorites**

... or fires, explosions  
in tornadoes eating network



**Management**

You must touch all redundant  
devices to make a change



**Messaging**

calling protocols  
spread contagion in the

**Management operations are a dominant trigger of correlated-failure network outages**

*[Evolve-or-die Sigcomm '16, Facebook '15, Microsoft '12]*

**Automation can make this better, but can often make it worse**

# | Drive defensively



Scripts are unsafe network automation.

Network management software should be written and operated like the production service it is

04

When you're done,  
you're just getting  
started

*and you don't know where this journey  
will take you*

# | No design survives contact with 10x growth



When you grow 10-15% a year, your designs will last 8 years

When you grow 40-50% a year, you get 5 years

When you grow 2-3x a year (pandemic serving), you have no time

Growth can be non-uniform, so stress cracks can appear earlier in different parts.



# | Build on the best aspects of each project



From Espresso, we expanded our network automation to every router in the network.



Our egress traffic engineering system can serve 15-20% more video for the same cost, while improving user experience by 2x fewer rebuffers and 2x higher video bitrate



Our globally connected network is now a key differentiating factor to allow Google Cloud customers to reach their users as well.



What will we get out of today's projects?

# 05

Watch what breaks, but  
also what works

*... just because it ain't broke doesn't mean  
you don't need to fix it*

# | Inspect the entire airplane

Don't just focus on the outages. Sometimes, systems that are working are slipping, but just not enough to cause trouble

This can be an early warning sign

Every system should have metrics it's measured against, and these metrics should be monitored

I AM WHEEL. HEAR ME SQUEAK.





| Thank you.

