Internet Centrality and its Impact on Routing

Geoff Huston AM
Cheof Scientist, APNiC
May 2022

Where does "centrality" come from?

- Let’s quickly look back a few years and the start of the Internet

Deregulation

- By the 1980's the telephone sector had exhausted our patience
  - They had become monopolies who were actively stifling competition
- "Deregulation" was the answer and the various national monopolies were broken up
- The idea was the competition in telephony would reduce prices for consumers

Competition

- The deregulation of the telephone industry occurred at much the same time as the rise of packet switched computer networks
- Packet switched networks were far more cost effective than voice-based circuit switched networks
  - Much of the cost in telephone networks was the time switch component
  - When you provide services to asynchronous applications the network platform is orders of magnitude cheaper

The Triumph of Competition

- The entire communications industry was broken apart and every element of the technology supply chain was opened up to competition
- And for a while this worked!
  - New data and computing providers competed with the former telco incumbents for market share: Cisco, Juniper, Intel, Microsoft, Amazon, Google
  - Consumer prices fell and consumers enthusiastically embraced the Internet environment
- Intense competition meant that providers had to:
  - Produce goods and services that consumers wanted
  - And do so at a price that was competitive with others in the market
- The consumer wins!

Deregulation and Markets

- The Internet was a poster child of deregulation of the telecommunications industry
- No more stifling telephone companies extracting monopoly rentals from consumers and supressing technology innovation
- We were meant to unleash market-driven competition upon the sector to:
  - Focus on what users want
  - To offer users real choice
  - To reward the most efficient and relevant providers
  - To quickly retire unwanted or inefficiently operated services

What happened since then?

- The last couple of decades has seen the decline of competition in this sector
- In place of “competition” we are seeing “centralization”

Today's Internet Centrality

- Linux completely dominates the platform OS space
- Chrome completely dominates browser space
- Mobile devices are either made by Apple and run iOS (30%), or run Android (70%).
- One quarter of the Internet's users send their DNS queries to Google's DNS open resolver service
- Google has such a massively dominant position across advanced technologies that everyone else is left trying to just follow and fill in the gaps – HTTP/3, QUIC, BBR, to name just a few
- Amazon had captured 50% of the retail e-commerce market in the US
- The entire world now finds itself relying on just one chip manufacturer!
  who is having trouble keeping up with orders!

Today's Internet Centrality

- We now have the problems of a monoculture with critical vulnerabilities
  - For example, 2021 outages by Fastly, Akamai and Facebook had global impact affecting all kinds of enterprises
  - A DDOS attack on DYN managed to cause a meltdown of the Internet for the entire eastern seaboard of the US a few years ago
- We are all relying on a very small number of service providers to be totally resistant to all forms of accidental and deliberate disruption

How did we get here?

- I thought that we were building a decentralised system with no critical orchestration components
  - No “Route Master 3000” in the middle controlling the routing space
  - No ”Name Czar” controlling all DNS resolution queries
  - No “Internet Service Controller” managing all Internet services
- All this was meant to be decentralised and adaptive
- We were meant to route around damage of any form

- Yet this is not exactly what we have today
- We’ve built something entirely different

Why are we doing this?

Why are we building Internet infrastructure that has:

- Limited diversity of supply
- Greater points of critical vulnerability
- Increased reliance on a limited set of service orchestration elements

What's driving change today?

Scale

“The only real problem is scaling. All others inherit from this.”

Mike O’Dell

For the past three decades we’ve been working as hard and as fast as we can to increase the size of the internet by a factor of a billion or so every decade!

That has been a task that has dominated our entire attention in the Internet

How have we responded?

- How has the Internet responded to these pressures of inexorable scaling in the Internet?

Building Bigger

Today's network architecture is based on **abundance**, not rationed scarcity

- Increasing **transmission capacity** by using photonic amplifiers, wavelength multiplexing and phase/amplitude/polarisation modulation for fibre cables
- Serving content and service transactions by distributing the load across many individual platforms through **server and content aggregation**
- The rise of high capacity mobile edge networks and mobile platforms add massive volumes to content delivery

Building Faster

- Reduce latency - stop pushing content and transactions across the network and instead replicate service and **serve from the edge**
- The rise of CDNs that serve (almost) all Internet content and services from massively scaled distributed delivery systems.
- “Packet Miles” to deliver content to users has shrunk – fewer miles means faster service!
- The development of mm wavelength cellular data systems (4G/5G) has resulted in a highly capable last mile access network with Gigabit capacity
- Applications are being re-engineered to meet faster response criteria

- Compressed interactions across shorter distances using higher capacity circuitry results in a much faster Internet

Building Better

If “better” means “more trustworthy” and “more privacy” then we are making progress at last!

- Encryption is close to ubiquitous in the world of web services
- TLS 1.3 is moving to seal up the last open TLS porthole, the SNI field, using ECH
- Oblivious DNS and Oblivious HTTP is moving to isolate knowledge of the querier from the name being queried
- The content, application, and platform sectors have all taken the privacy agenda up with enthusiasm, to the extent that whether networks are trustable or not doesn’t matter any more – all common shared network infrastructure is uniformly treated as untrustable!
- And if you can’t trust it then avoid it (where possible)!

Building Cheaper

- We are living in a world of abundant comms and computing capacity
- And working in an industry when there are significant economies of scale
- And being largely funded by a small number of enterprises capitalising through advertising a collective asset that is infeasible to capitalise individually
- The result is that a former luxury service accessible to just a few has been transformed into an affordable, fast, mass-market commodity service available to all
  - but provided by a small clique of providers

So it's all good?

Right?

So it's all good?

Or maybe not.

How to answer this question?

- Let’s dive in a bit deeper and look at this through the lens of the “shape” of the Internet
- How has the Internet changed to allow us to build bigger, faster, better and cheaper?
- We have changed the way the Internet handles routing

Scale and Routing

"The most complicated computation ever attempted by mankind is the global distributed routing algorithm that runs the Internet.

In fact, if anybody thought about it very hard, before we started, they would've been too scared to try.

Ah, because it runs in near real-time, it's an online algorithm, it runs on a multimillion node multicomputer, of an arbitrary topology, built by lots of people who have never met each other. Right?

And, it's a very very complex computation because it's piecewise constructive, there is a lot of local consistency constraints, there is a bunch of global correctness criteria that are occasionally satisfied, and yet the thing mostly works.

Which is astounding, when you actually look at what's going on."

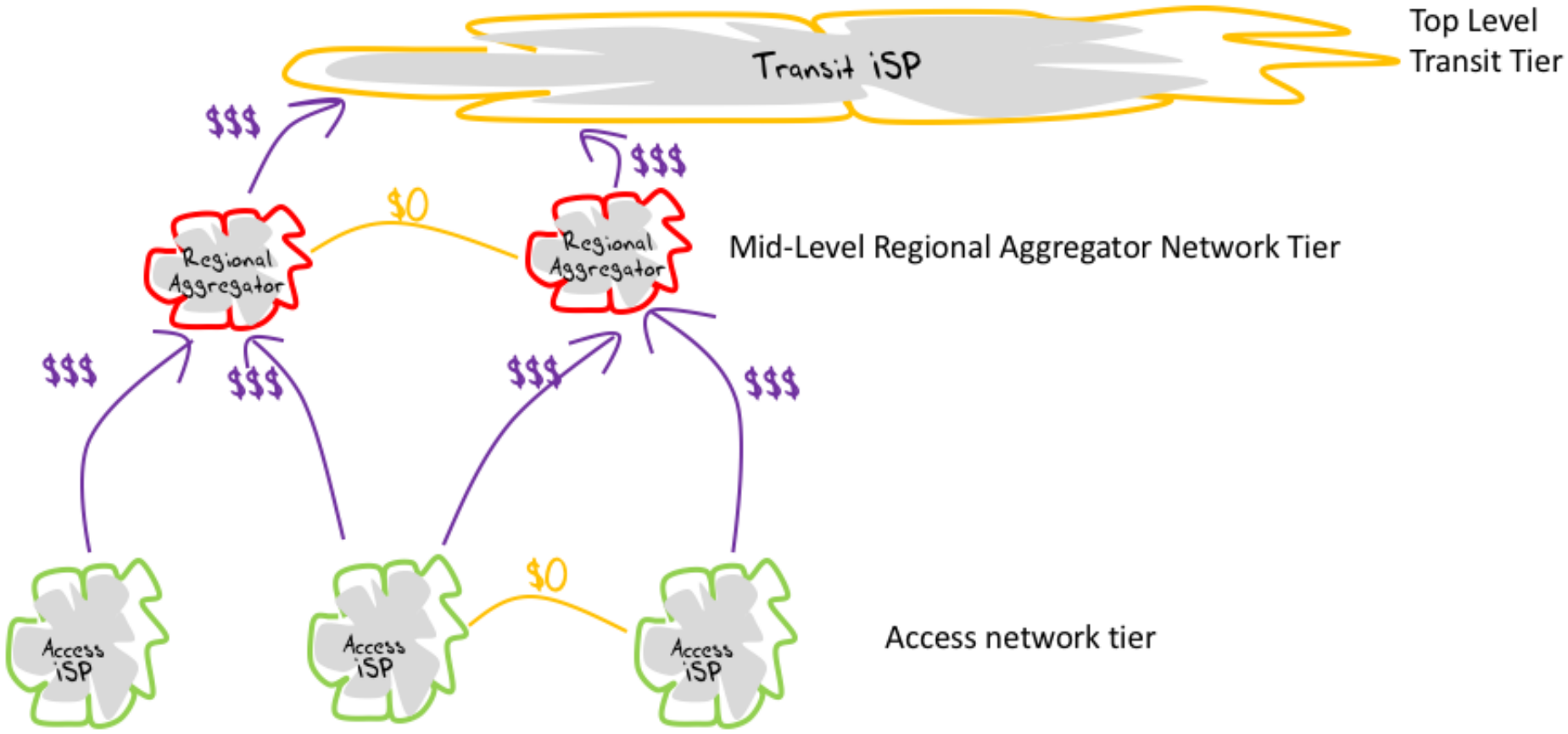
Also from Mike O'Dell, 2000 (http://www.dtc.umn.edu/~odlyzko/isources/odell-transcript.txt)

Avoiding Routing

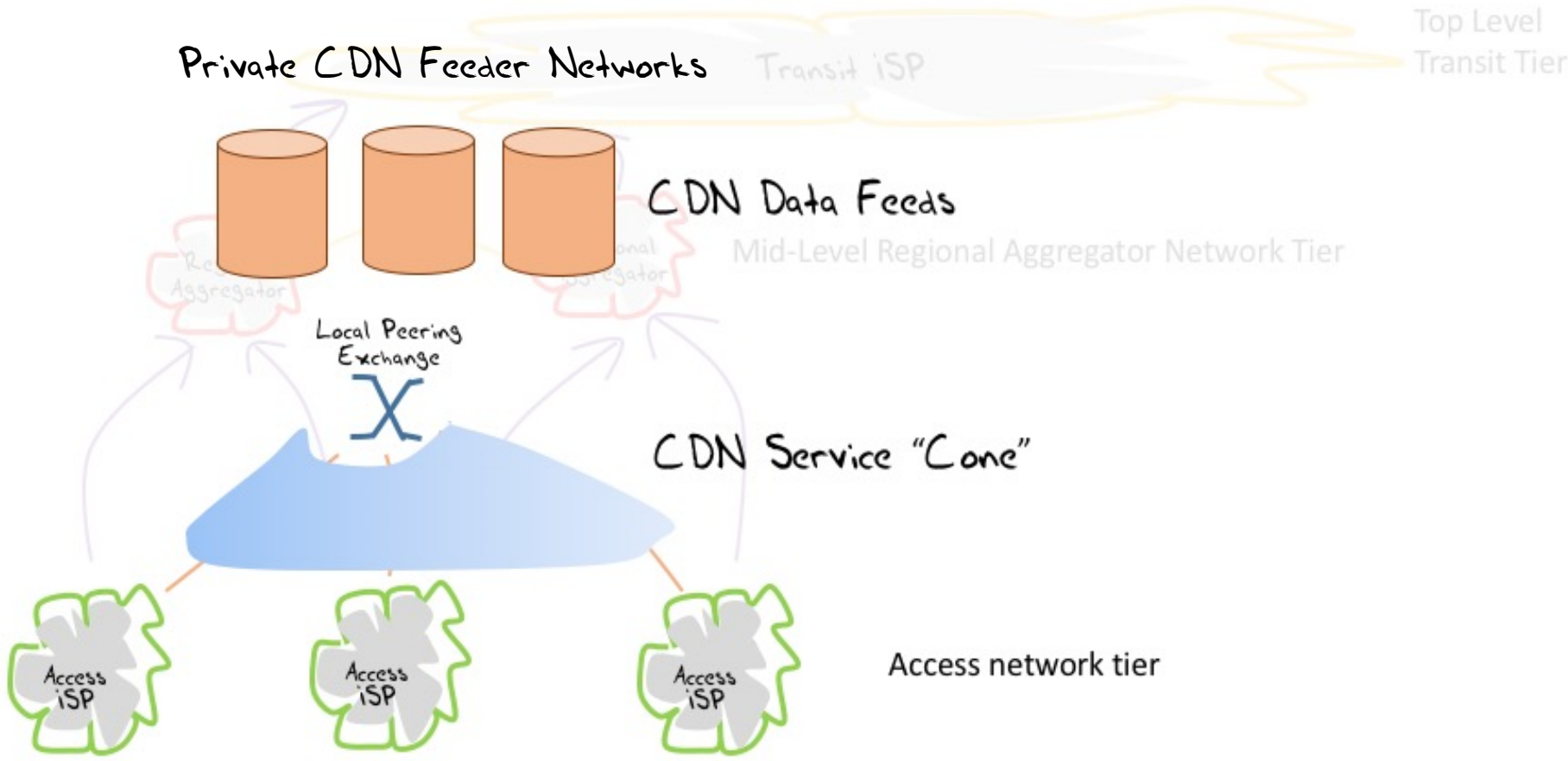
Pushing EVERYTHING out of the network and over to the edge!

- Transmission infrastructure is becoming an abundant commodity
  - Sharing technology (packet multiplexing) is decreasingly relevant is every large service platform migrates its own dedicated distribution system
- We have so much network and computing that we no longer have to bring consumers to service delivery points - instead, we are shifting services towards consumers and using private network slices to replicate service delivery from densely deployed data centres
- With so much computing and storage available the application is becoming the service, rather than just a window to a remotely operated service operated on expensive computing and storage platforms

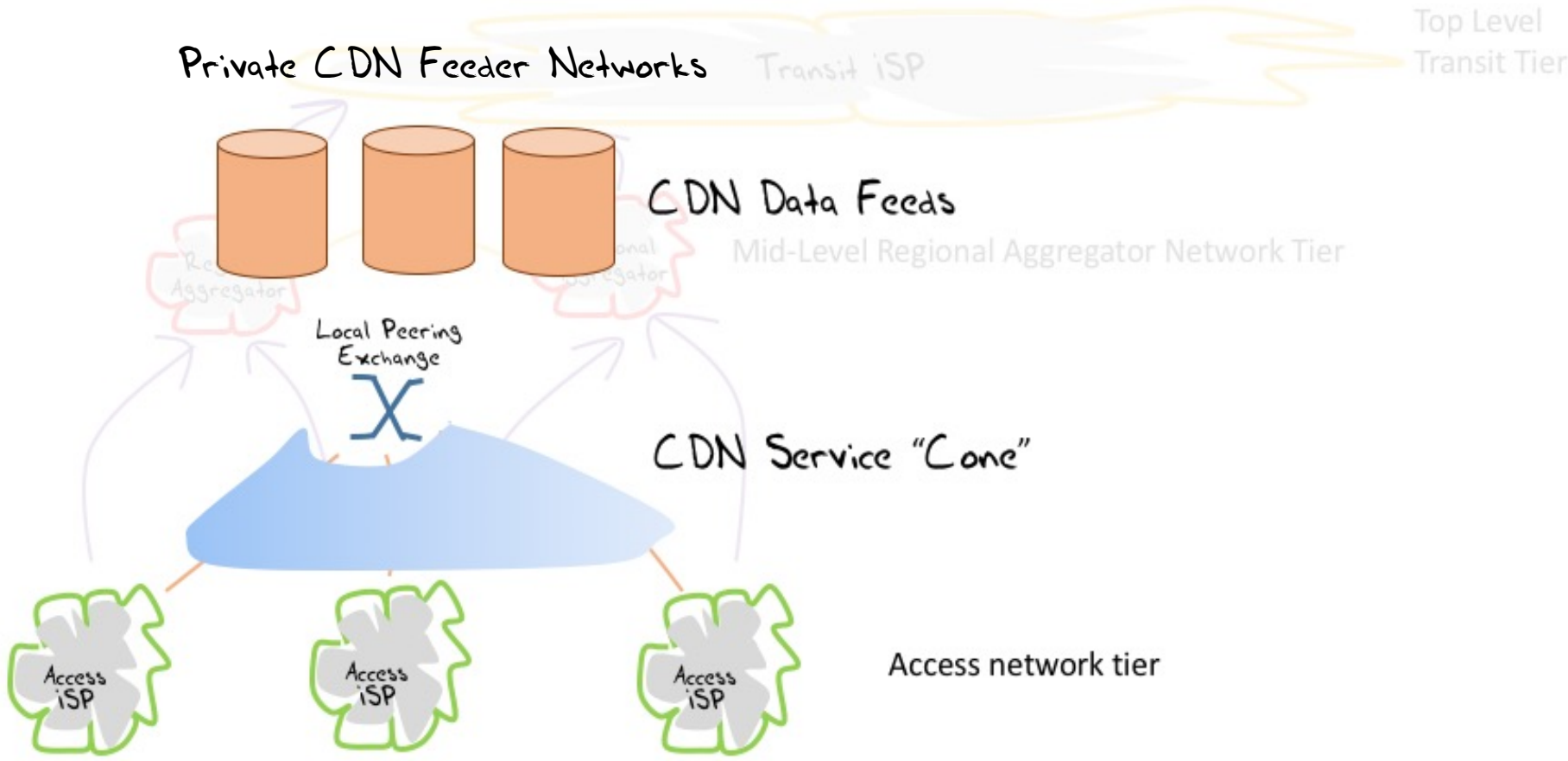
The 1990's Internet



Today's Internet Architecture



Today's Internet Architecture



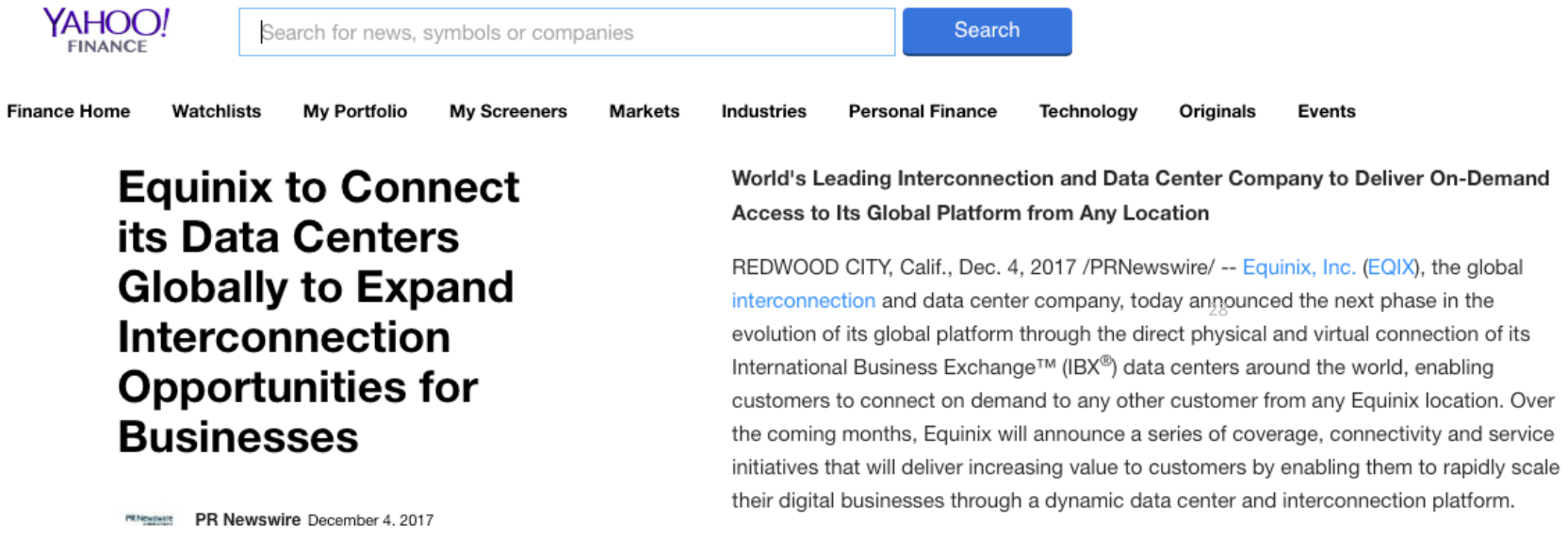
Who needs Transit?

- If users don't send packets to users any more…
- If content is now delivered via CDNs to users via discrete service cones…
- If there is no universal service obligation…

Then why do we still need Transit Service providers?

Closed Transit?

We see the CDN systems reserve a carriage resource through dedicated bandwidth / wavelength / cable purchase and effectively bypass the open IP carriage infrastructure



Transit?

Once the CDN caches sit “inside” the Edge NAT of the Access ISP then the entire wide area network becomes a marginal activity compared to the value of the content feeds!

Transit Routing

- How much data is delivered from the local Data Centre through Access network to the end device?
- How much data is delivered across just 1 AS hop?
- Who *needs* to route with a full routing table?
  - From the perspective of the edge are we just making a routing decision within local datacentre to select services points hosted by Amazon, Google, Cloudflare, Akamai, Fastly, Microsoft and Apple?
- What's the cost and benefit of a transit and a full routing table any more?
  - Who is willing to spend 80% of their effort and cost providing a service for less than 1% of their traffic and less than 1% of their revenue?

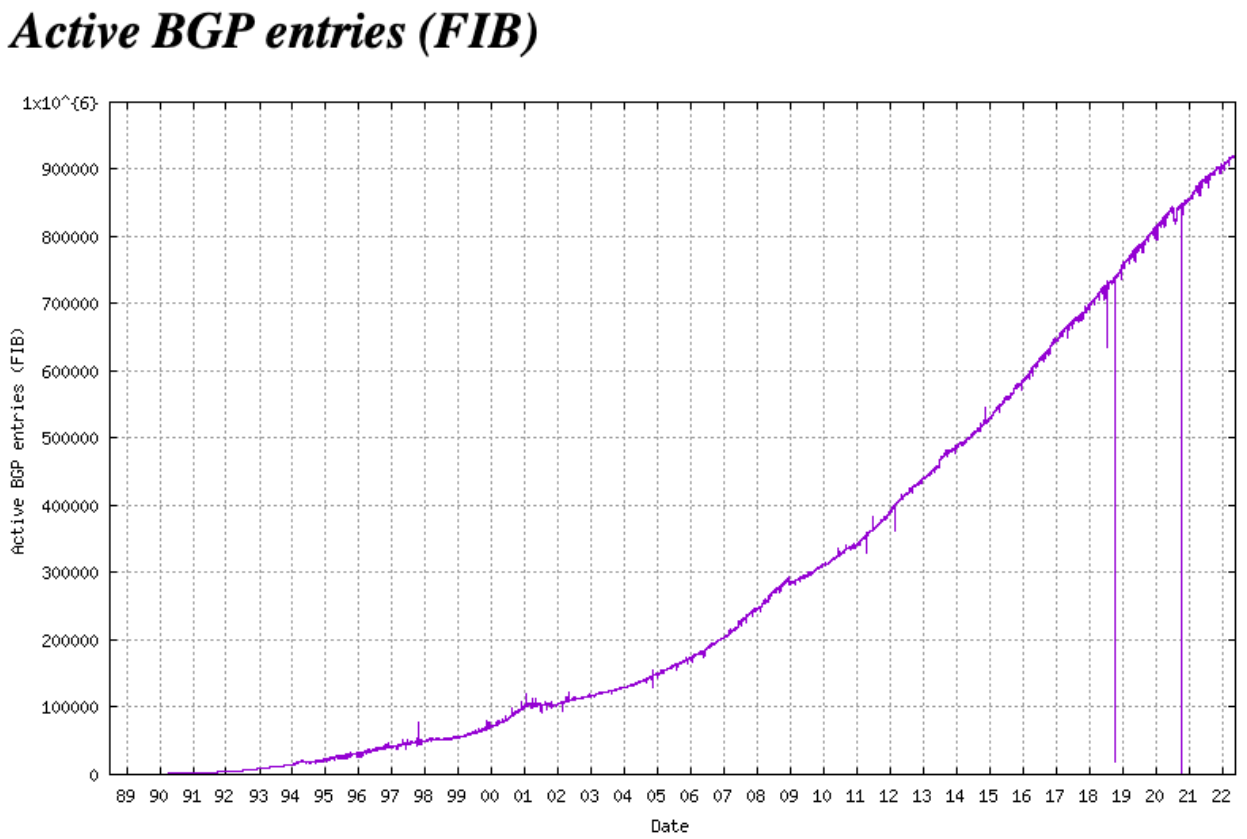
Internet Names and Addresses?

If the Internet is just a collection of discrete CDN service ‘cones’ then why should we expect end users to pay for the maintenance of:

- A global address plan?
- A global name system?
- A single global routing system?
- A single global network?

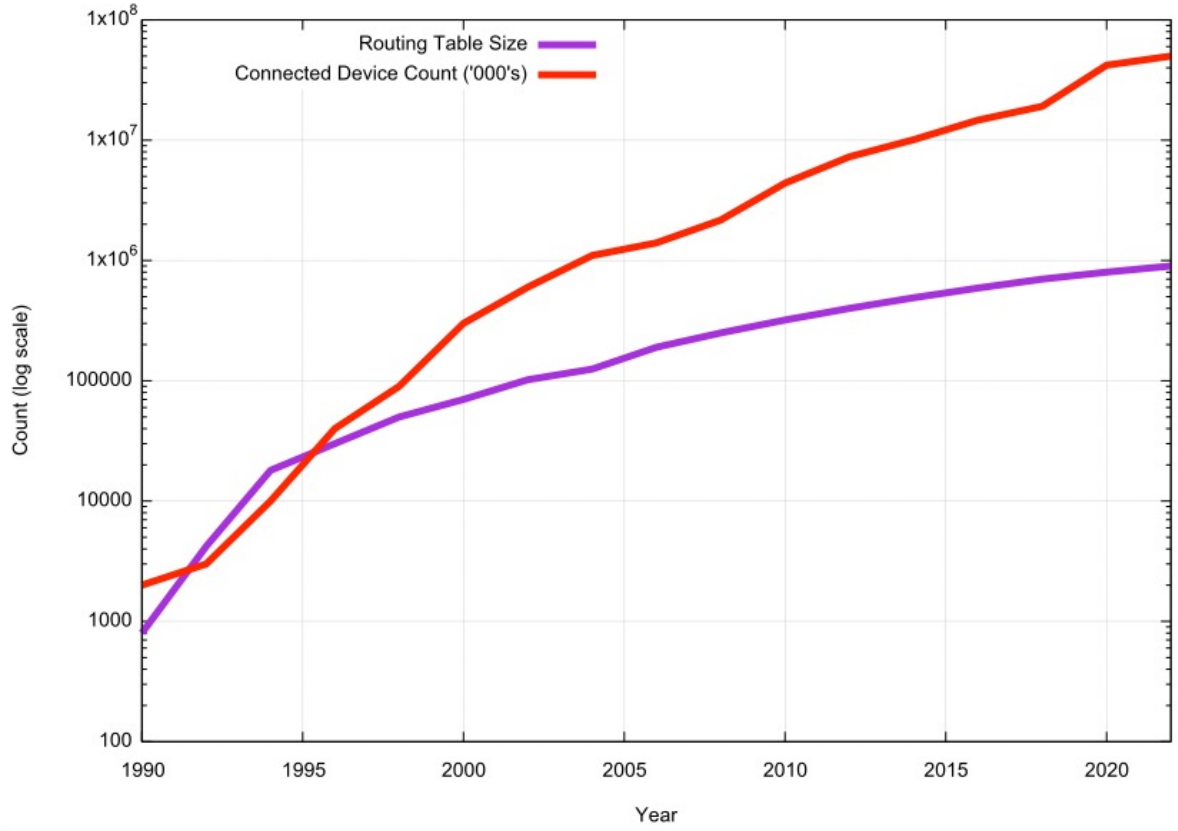
The Routing System

- If we don’t need global address plan any more then why does the routing space continue to grow?



Routing Trends

- As we push the delivery content and service off the edge of the network we are reducing the scaling pressure on the routing environment
- The connected device count is growing at a faster rate than the routing system



Exactly where are we now?

- We started this journey building a telephone network for computers to communicate between each other
- But we've moved a long way away from that model:
  - one-way content distribution lies at the core of today's Internet
  - This content distribution role is an enterprise service framework rather than a public carriage service
  - The internal parts of the carriage network are now being privatized and removed from public oversight

Commoditization and Scaling

- We should expected all this to occur
- The pressures to keep on reducing the unit cost of service delivery while also allowing the Internet to scale to support more data intensive services has meant that we can only grow by commoditizing the lower layers of Internet infrastructure
- In commodity markets, economies of scale are paramount
- Only a small number of large providers are viable - centrality
- This has implications in the architecture of the network itself

It's not just the Death of Transit

It's the rethinking of the entire network as a common shared glue

- Service provisioning sits within cloud providers and distributed data centres
- Edge computers are now acting as televisions into the clouded world of data
- The distinction between personal and public data realms is disappearing into the realm of corporately owned private data empires
- And the barriers to entry for new players get increasingly larger as the platform incumbents bolster their role to replace a single common network with a small number of such platforms in a cartel-like arrangement

  (Today's version of "centrality")

Do shared networks matter any more?

- We have increasingly stripped out network-centric functionality in our search for lower cost, higher speed, and better agility
- We are pushing functions out to the edge and ultimately off “the network” altogether and what is left is just dumb pipes
- What defines “the Internet”?
  - A common protocol, common protocol address pool, and a common routing system?

  or
  - A disparate collection of services that share common referential mechanisms?

Thanks!