Internet Evolution (and IPv6)

Geoff Huston AM

Chief Scientist, APNIC

The Internet used to be simple…

1980's:

- The network was the transmission fabric for computers
- It was just a packet transmission facility
- Every other function was performed by attached mainframe computers

"dumb" network, "smart" devices

Then we went client/server

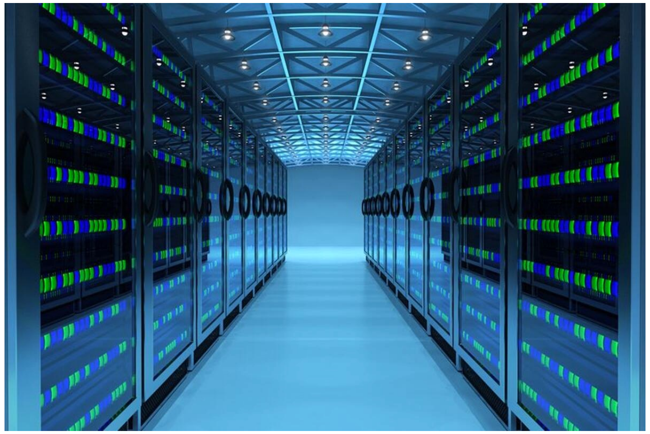
1990's:

- The rise of the Personal Computer as the "customer's computer"
- We started to make a distinction between "customers" and "network"
  - The naming system was pulled into the network
  - The routing system was pulled into the network
  - Messaging, content and services were pulled into the network
- We created the asymmetric client/server network architecture for the Internet

Internet Infrastructure of 2000

Rapid expansion of network infrastructure in many directions:

- Exchanges, Peering Points and Gateways
- Transit and Traffic Engineering
- Data Centres and Service "Farms"
- Quality of Service Engineering
- MPLS, VPNs and related network segmentation approaches
- Mobility Support – Mobile Networks
- Customer Access Networks
- Content Distribution Networks



Aren't these all "different" networks?

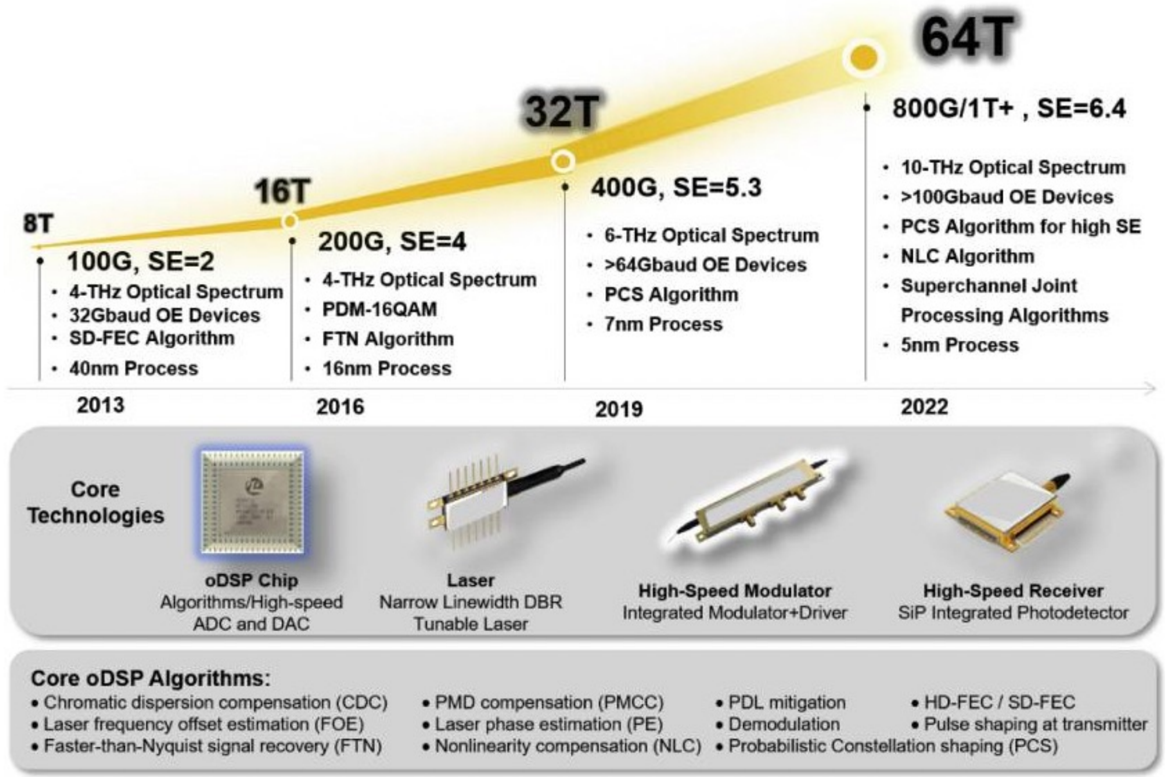
- Well, yes they are
- The true genius of the Internet was to separate the service environment from the transmission technology
  - Each time we invented a new comms technology we could just “map” the Internet onto it
  - This preserved the value of the investment in “the Internet” across successive generations of comms technologies

What about the coming decades?

- The seeds of the dominant factors of the future environment are probably with us today
- The problem is that a lot of other seeds are here as well, and sifting out the significant from the merely distracting is the challenge
- So with that in mind lets work out the big drivers in today's environment…

Abundant Capacity

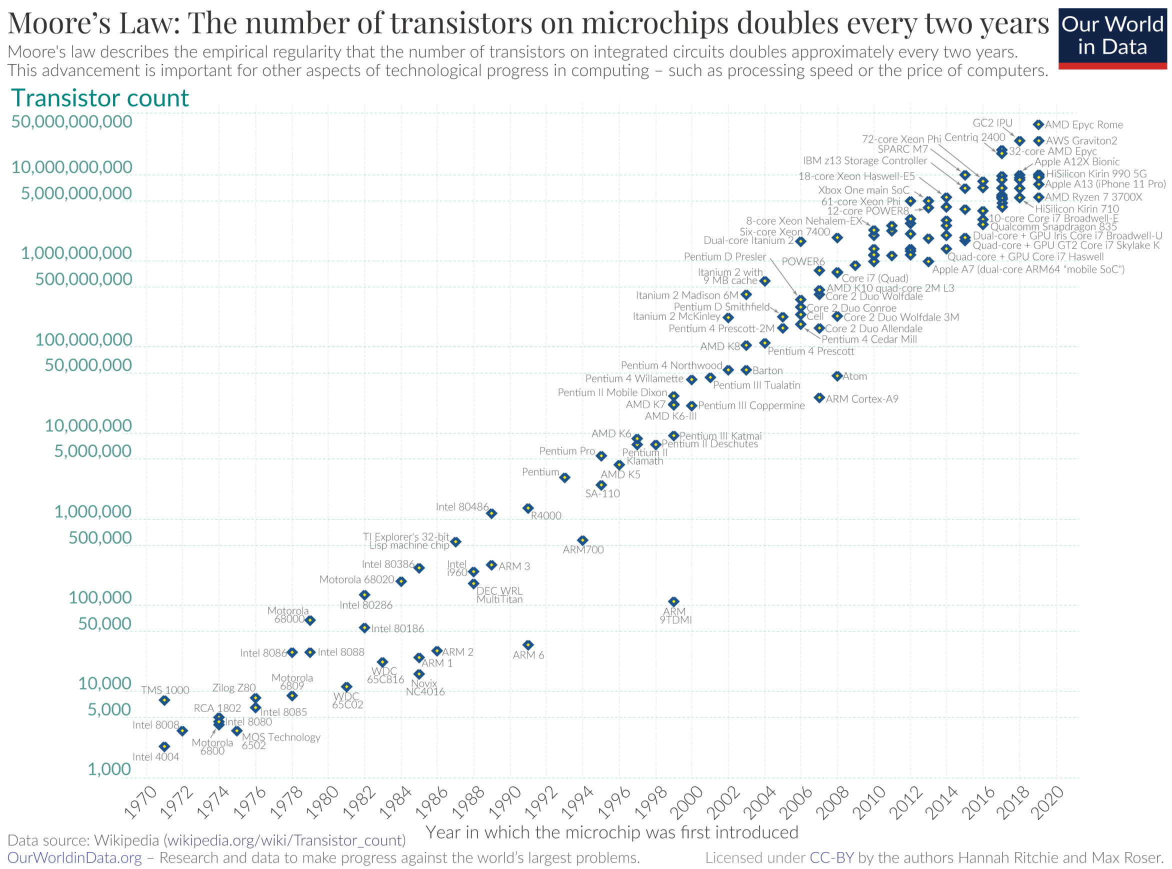
Fibre cables continue to deliver massive capacity increases within relatively constant unit cost of deployment



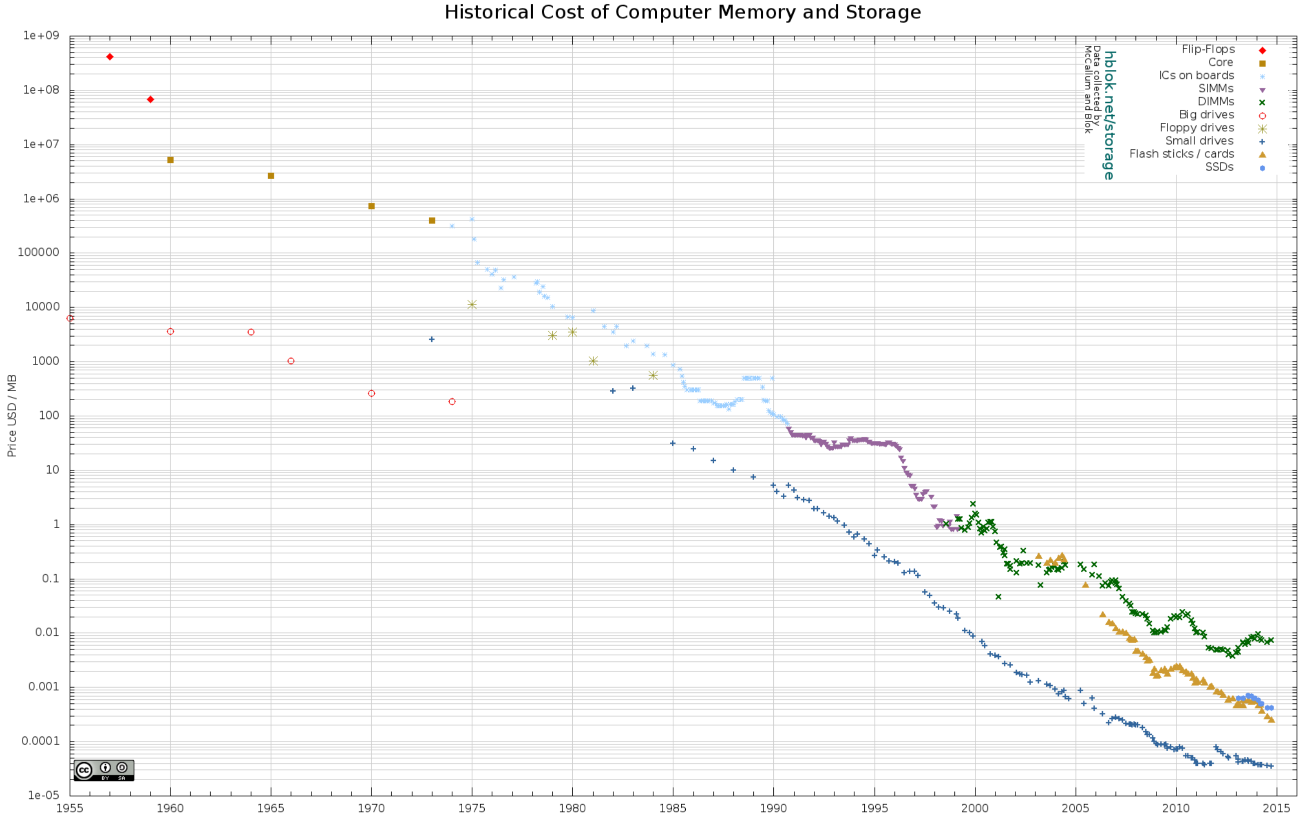
https://www.ncbi.nlm.nih.gov/

(That 2022 number is probably low – at the end of 2022 we can pull 2.2T per lambda with a 190Gbd signal rate, giving a fibre capacity of 105T)

Abundant Compute Power



Abundant Storage



http://aiimpacts.org/wp-content/uploads/2015/07/storage_memory_prices_large-_hblok.net_.png

What's driving change today?

- From scarcity to abundance!
- For many years the demand for communications services outstripped available capacity
- We used price as distribution function to moderate demand to match available capacity
- But this is no longer the case – available capacity in today's communications domain far outpaces demand

How can we use this abundance?

- By changing the communications provisioning model from ***on demand*** to ***just in case***
- Instead of using the network to respond to users by delivering services *on demand* we've changed the service model to provision services close to the edge just in case the user requests the service
- With this change we've been able to eliminate the factors of *distance* from the network and most network transactions occur over shorter network spans
- What does a *shorter* network enable?

Bigger

- 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

- To manage this massive load shift we've stopped pushing content and transactions across the network and instead **we serve from the edge**

Faster

- Reduce latency - stop pushing content and transactions across the network and instead **serve from the edge**
- The rise of CDNs serve (almost) all Internet content and services from massively scaled distributed delivery systems.

- The “Packet Miles” to deliver content to users has shrunk - that’s faster!
- The development of high frequency 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

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 capitalising a collective asset that is infeasible to capitalise individually – the advertisement market
- The result is that a former luxury service accessible to just a few has been transformed into an affordable mass-market commodity service available to all

So it's all good!

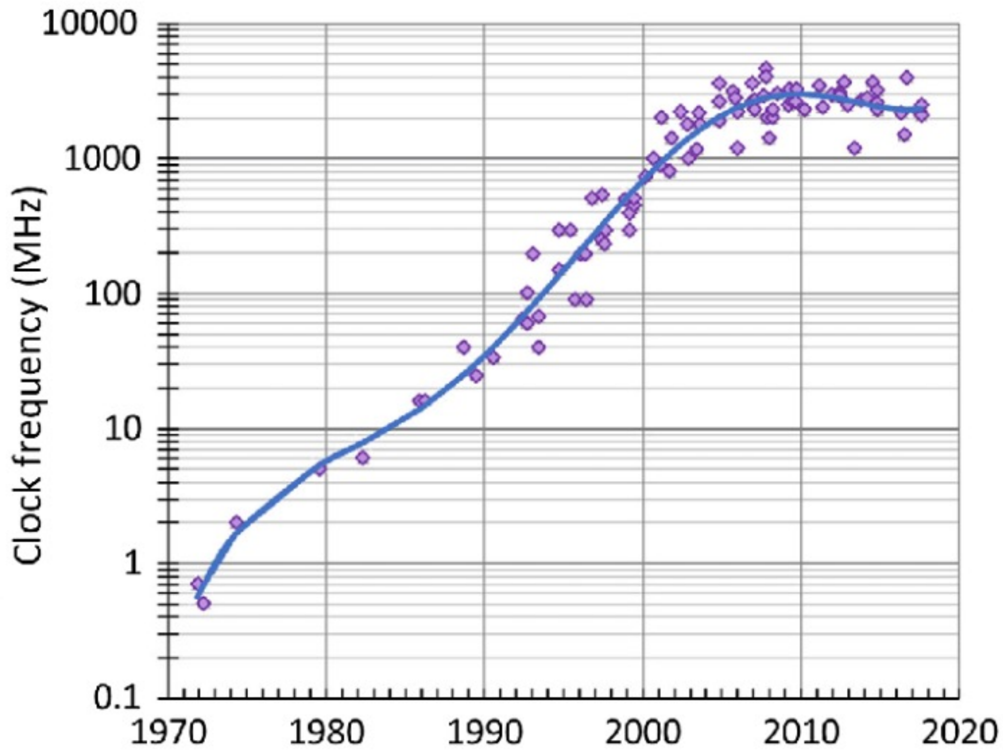
Right?

Not Quite

Processor clock speeds have topped out over the past decade

While the network growth trends continue to scale at an exponential rate, silicon-based processing capacity is now growing at a linear capacity at best

Why should we be concerned about this?

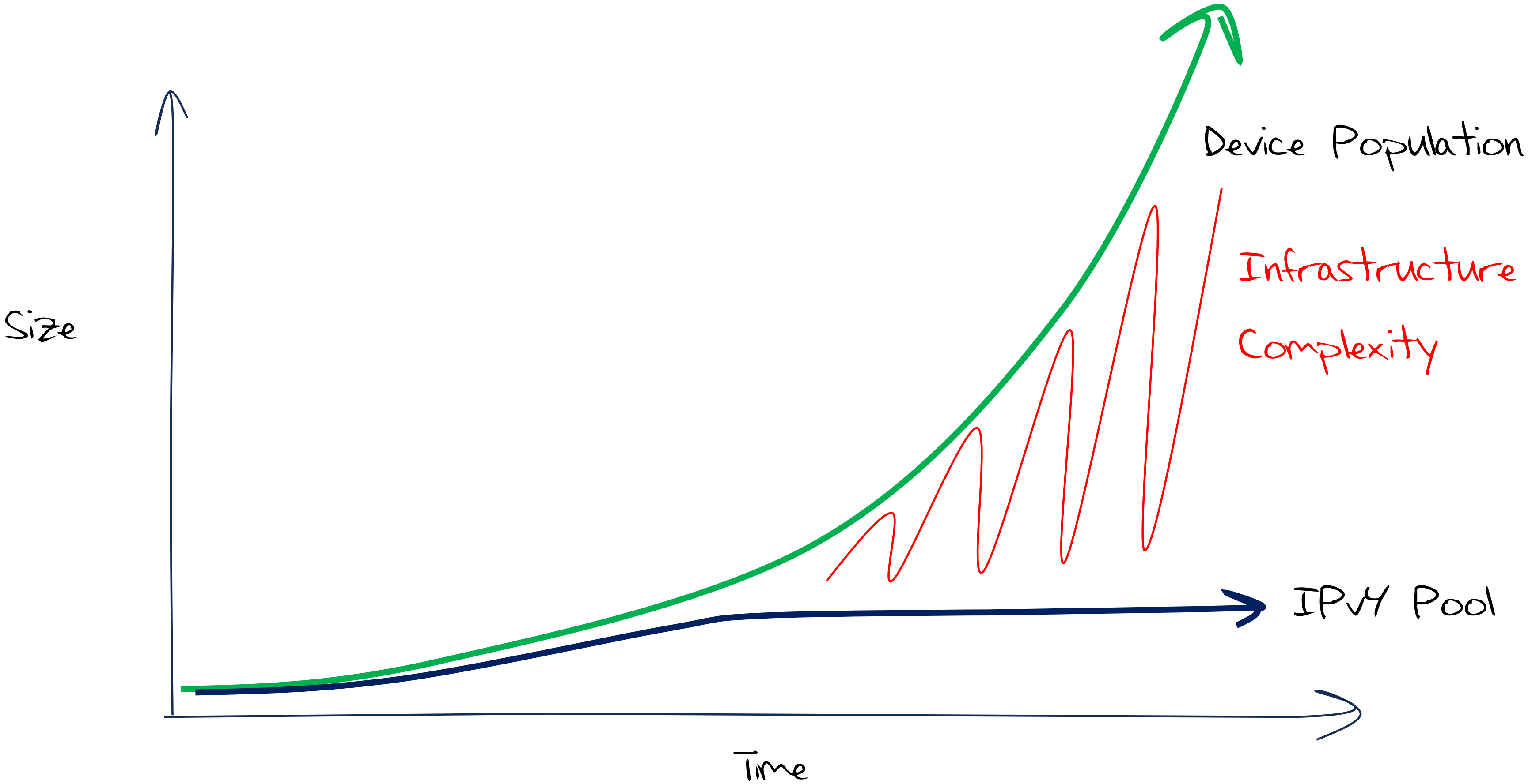


Processor Clock Frequency Trend Data

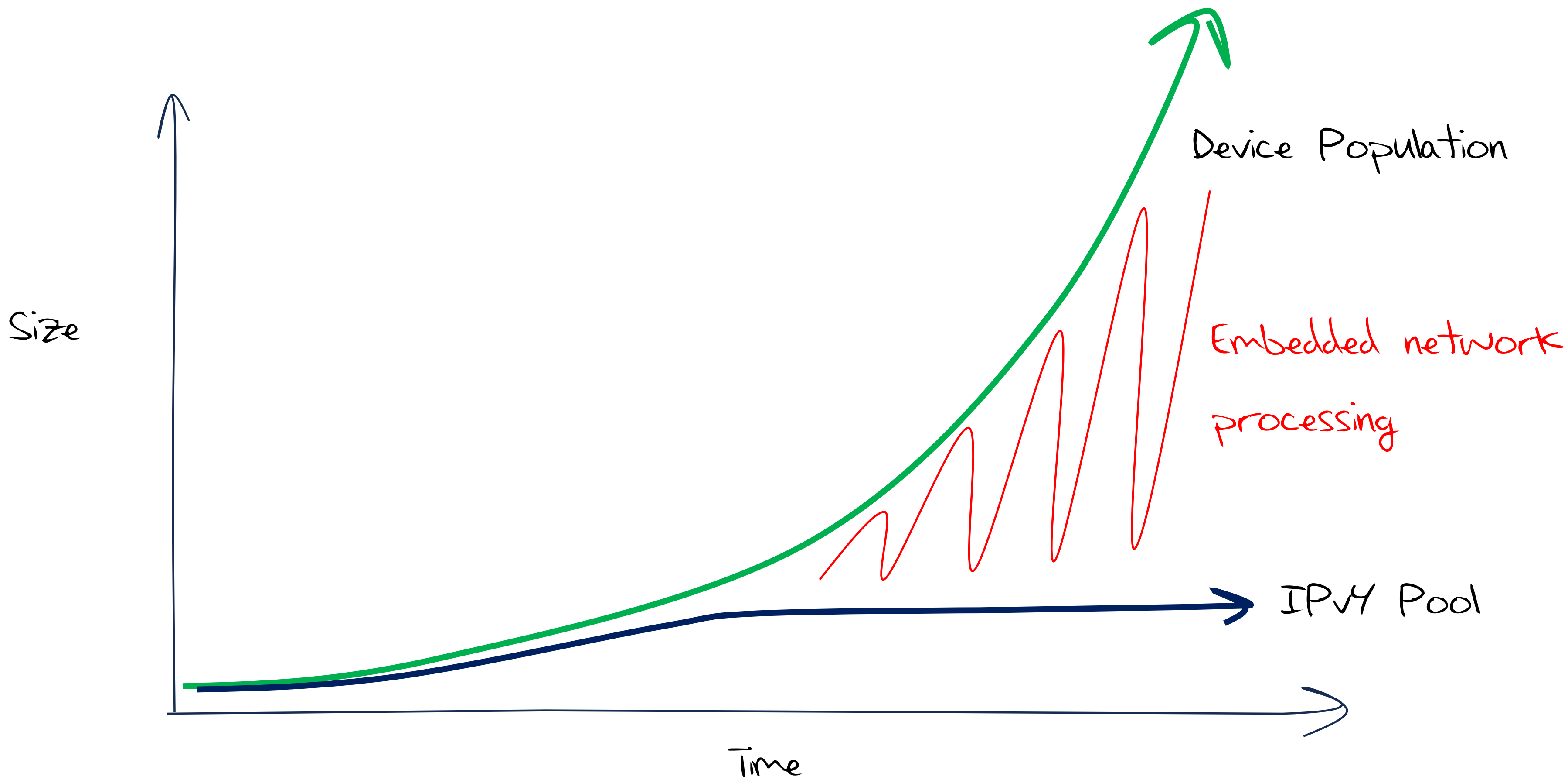
Internet Scaling

- To make up the shortfall in IPv4 addressing we've adding greater processing capability into the network's infrastructure
  - Network Address Translation, dynamic naming and content steerage
  - Replacing static data with on-demand processing
- This approach is viable in the long term only if we can scale processing efficiency in line with demand growth
- But if processing capability is not scaling then we have a problem …

Internet Scaling



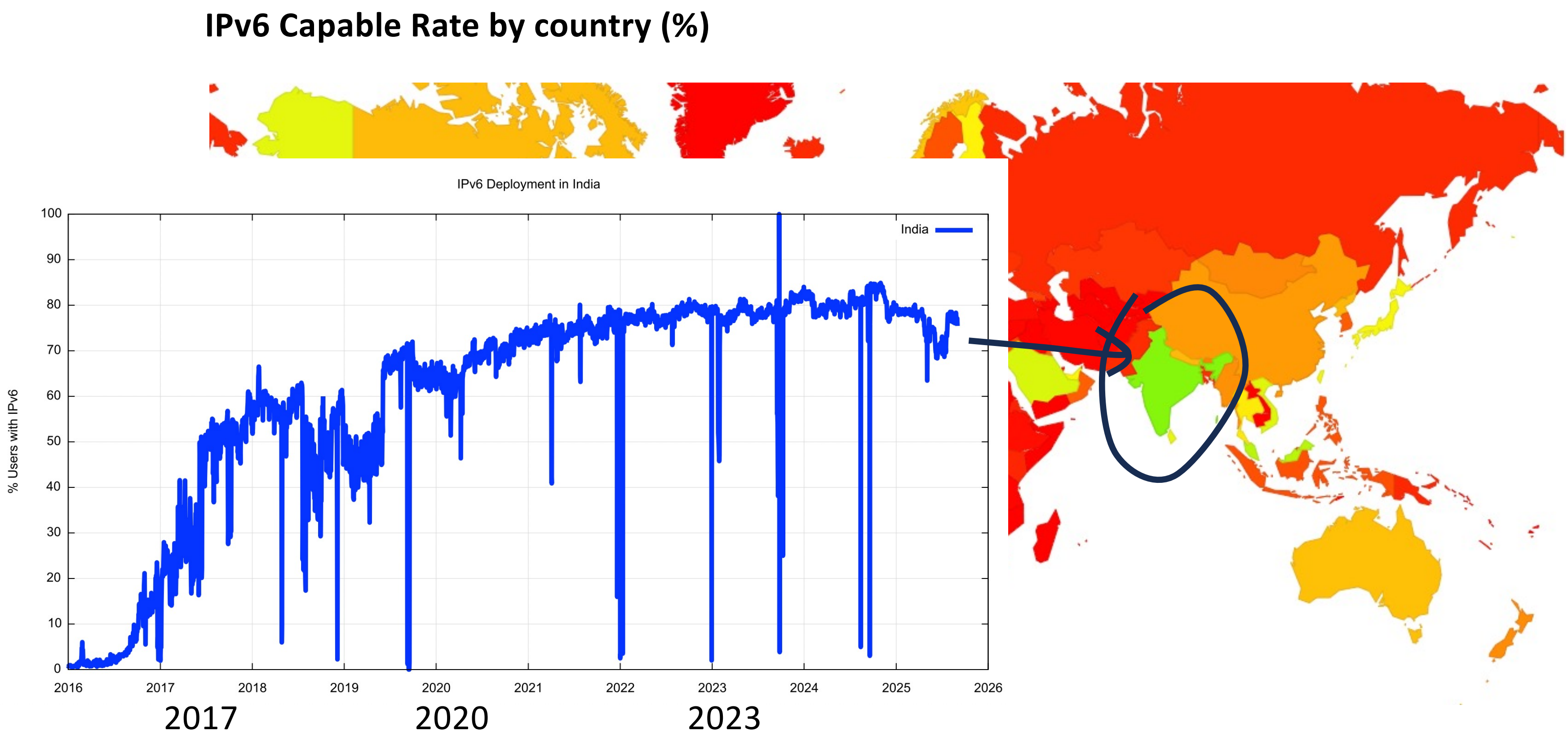
Internet Scaling



Internet Scaling

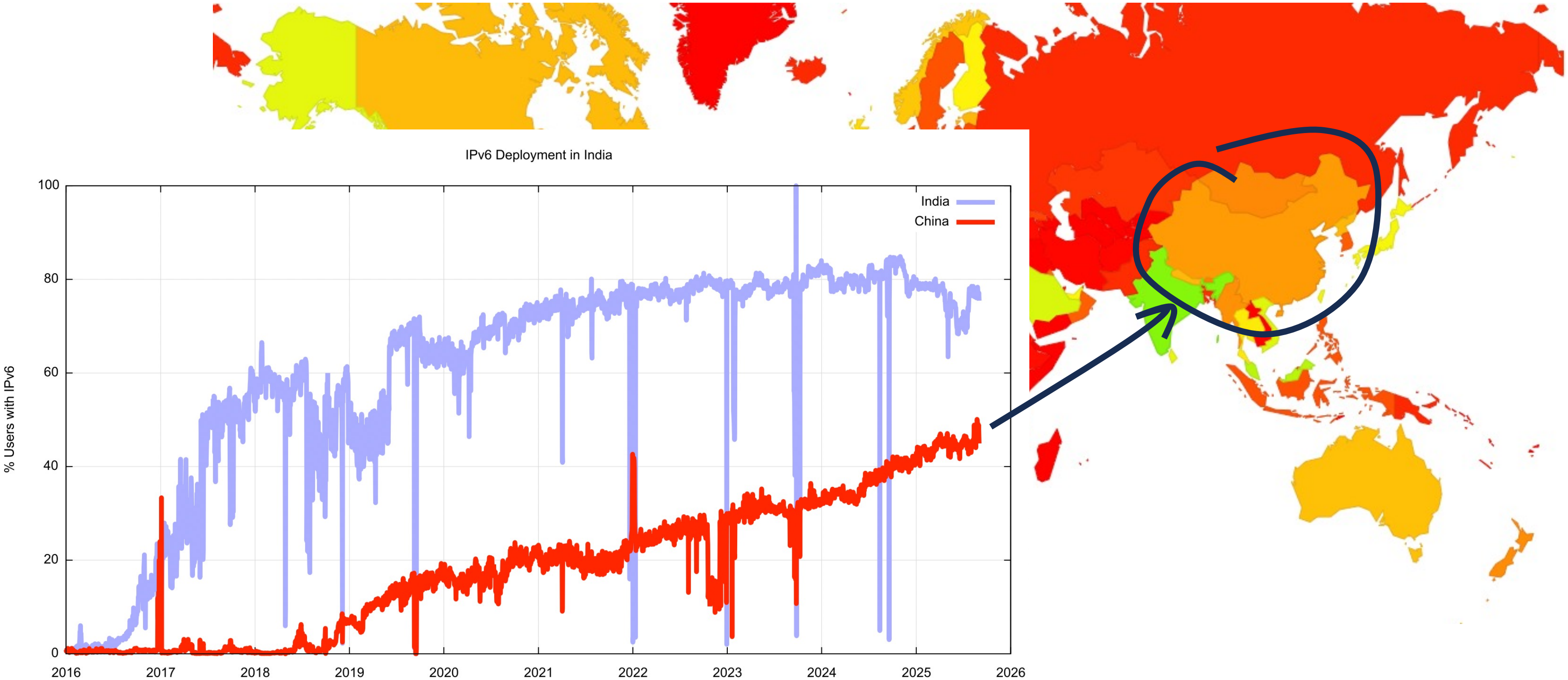
- If we cannot scale network complexity, then we need to reduce the complexity burden within the network
- But IPv4 exhaustion gives us little leeway to reduce network complexity
- Which means that if you want to:
  - Deploy digital services at scale
  - Contain cost escalation to keep the service affordable
  - Improve network robustness and security
- Then you have few choices left other than to reduce the network complexity burden
- Deploying IPv6 is one obvious response

Which might help to explain India's dramatic move to IPv6



And China's IPv6 efforts

IPv6 Capable Rate by country (%)



Longer Term Evolution

Pushing EVERYTHING out of the network and over to applications

- Leave the network's role to simple end-to-end transmission at scale
- Push service mediation roles out of the network and bring services towards consumers, using content distribution frameworks to distribute replicated servers and services
- T**he application is becoming the service**, rather than just a window to a remotely operated service

Evolutionary Shifts

The key innovation of the Internet in the 1980's was to push functionality out of the networks and into the connected hosts at the edge

- A simpler network allowed the network to scale at lower cost
- And scaling at the edge was a case of replication

We tied this together with a coherent address architecture

Evolutionary Shifts

Today we are moving away from host-centric services to application-centric services, pushing services and functions away from hosts and platforms into distributed shared state at the application level

- Services are defined by reference in a common name space
- Addresses are just tokens used to guide packets through the underlying connectivity mesh

We tie this together with a coherent name space – the DNS

Longer Term Evolution

- The infrastructure for applications and security is transitioning into the name space
  - The identity space and associated authenticity credentials are now associated with service names, not IP addresses
  - DNS services now underpin and define the Internet environment
  - IP addresses are still topology tokens used to for packet forwarding – but the association between a service and an IP address token is dynamically dcefined by the DNS.
  - This has removed the pressure from the transition to IPv6, as the DNS masks over the dual stack nature of the underlying forwarding network

Where does IPv6 fit?

- IPv6 allows us to move on from the issues of scaling the underlying connectivity fabric of the network with an inadequate pool of IPv4 address tokens
- Today's need is to scale the service environment so that we can meet the scale and capacity of service delivery requirements that exceed individual platform capabilities
- This **service level scaling** challenge is what should absorb our attention in the coming decade or more

Thanks!