

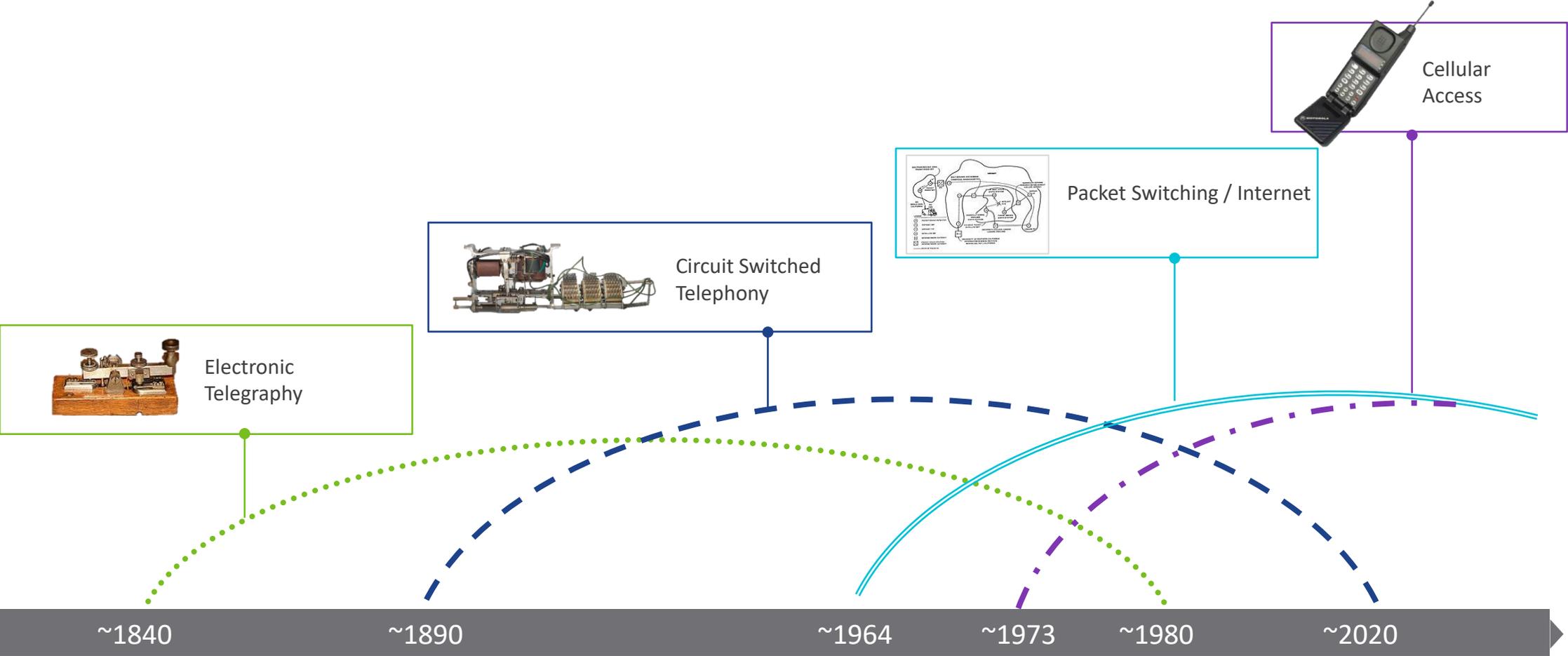
# Surprise-Inspired Networking

David Tennenhouse  
*Free Agent*

June 2022

# Disclaimers

# We innovate AND build long-lived stuff

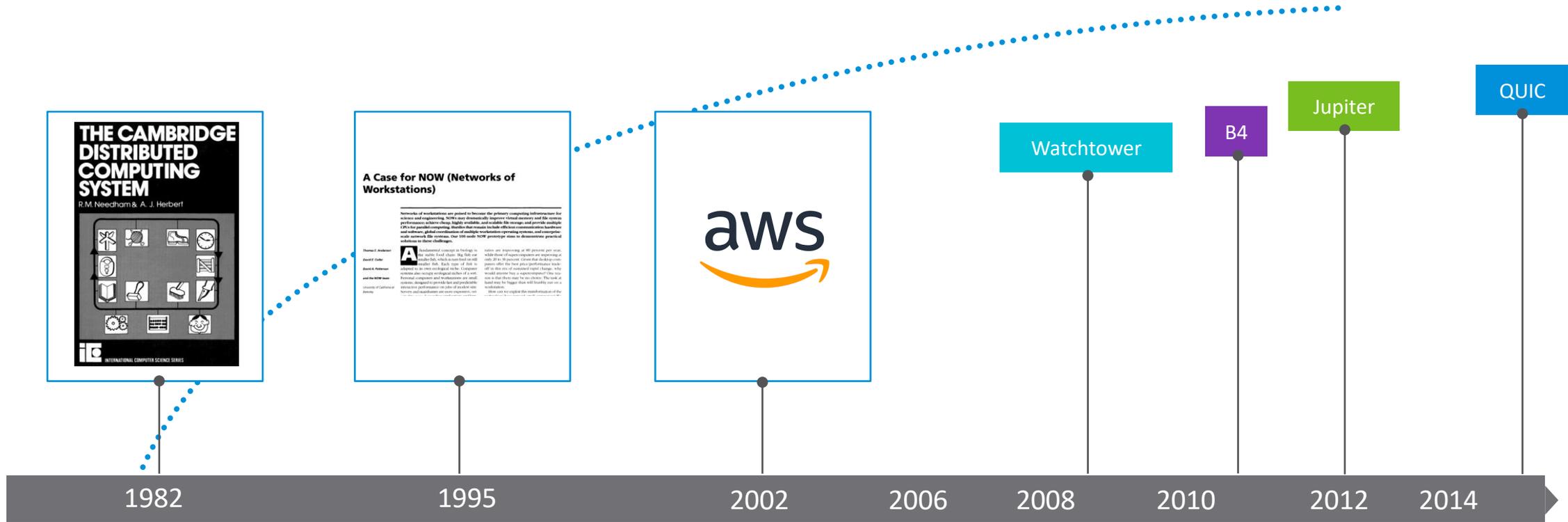


## Tools of our Trade

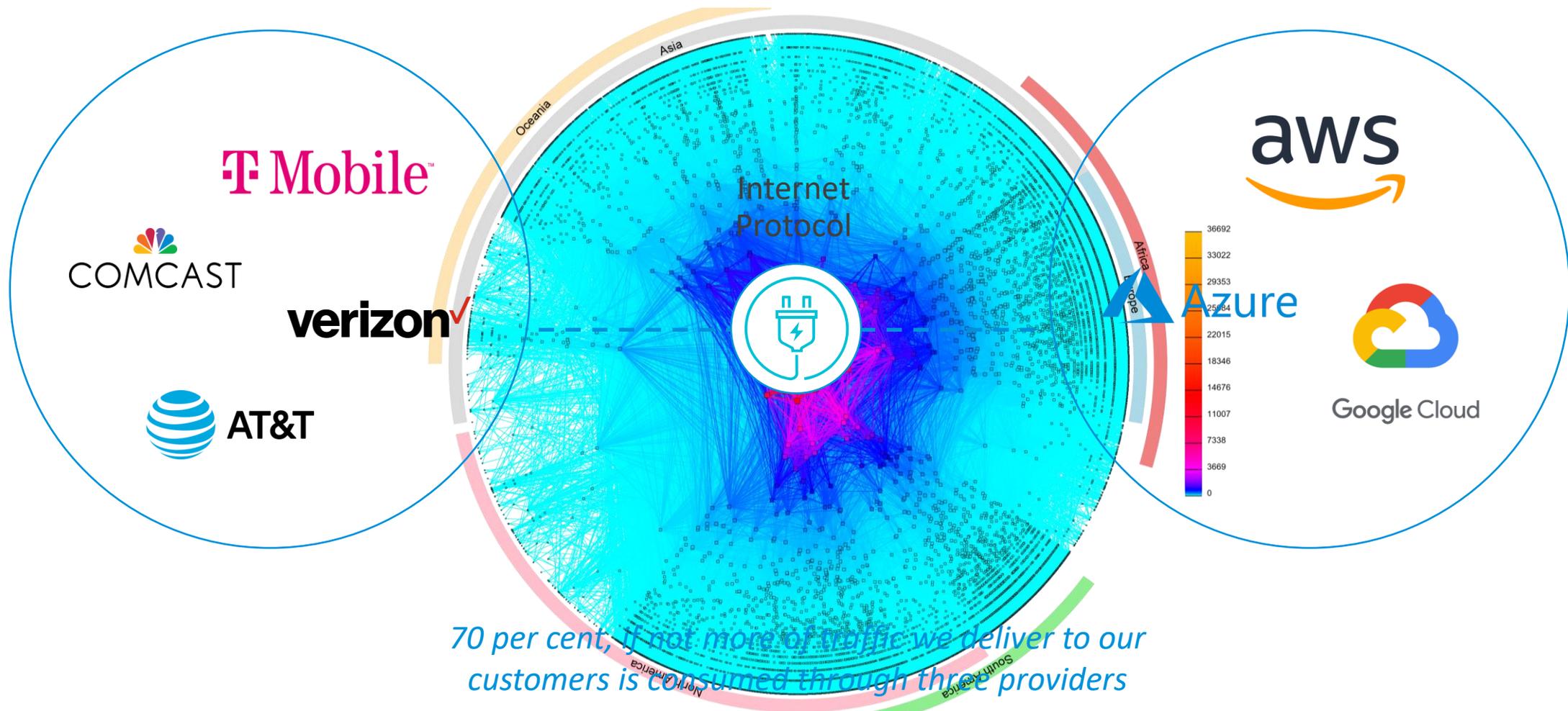
- Computation in the network
- Storage in the network
- Information Theory

# A lot has changed in the past 10 years

## ~~End-to-End Communication~~ Edge-to-Cloud Communication

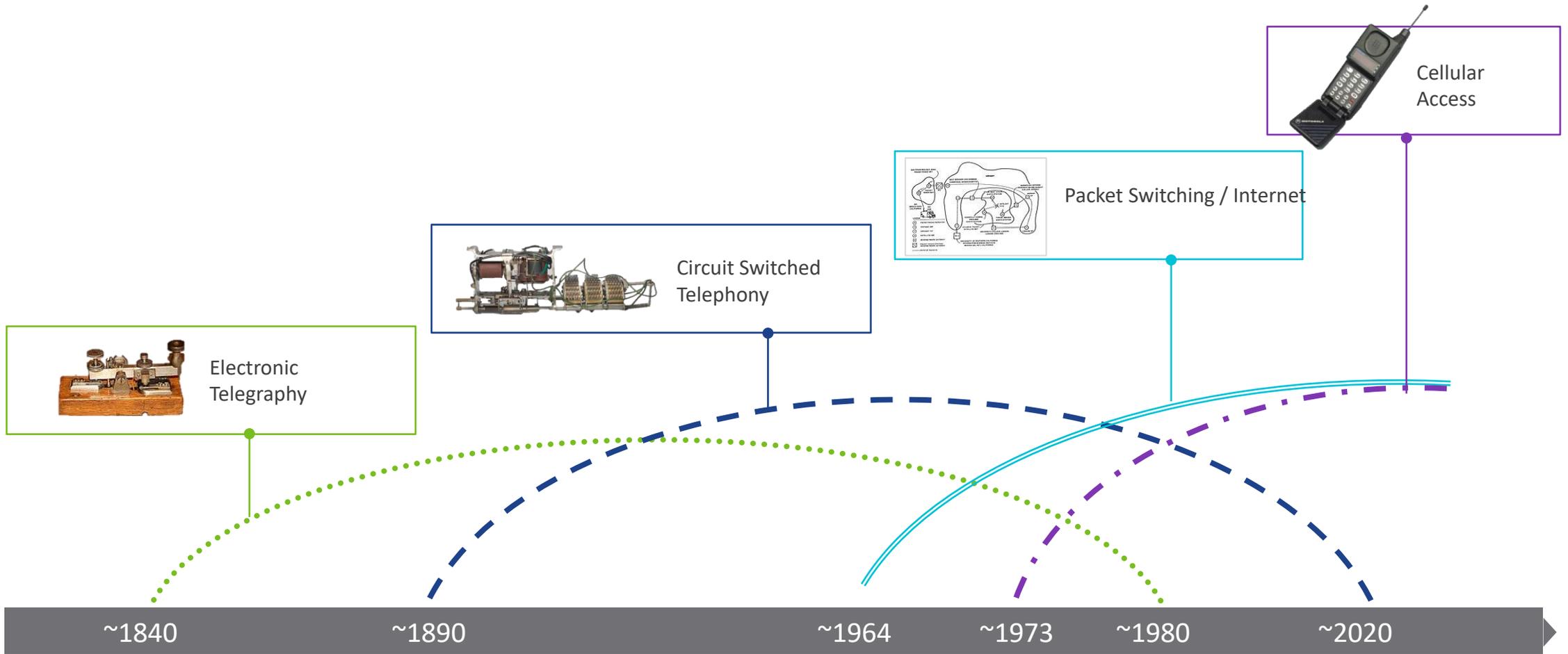


# Is the Future Internet a Global Network ... – or a point-to-point *connector*?



— Howard Watson, CTO, BT Group plc

# So...what's next?





# Surprise

The *information content* (also called the *surprisal*) of an event  $E$  is...

$$I(E) = \log_2 (1/p(E))$$

*The basic idea of information theory is that the ‘informational value’ of a communicated message depends on the degree to which the content of the message is surprising\**

## Since Only *Surprise* Matters...

- **Simplistic Interpretation:** Delivery of *old* content has no value
- **Realistic Interpretation:**
  - Be thoughtful about how/where *old* content is combined with *new* information
  - The value comes from the fusion of real-time and opportunistic computing

### ***Semantic Surprise:***

- Evaluate *surprisal* through the lens of the application(s), not just the channels
- Move beyond using information theory at the link layer

# Context: What Will The Future Applications Be?

## They will be Proactive

Machine-to-Cloud (M2C) systems that bridge the physical and virtual worlds

... and that *anticipate our needs and act on our behalf*

- Examples:
  - Augmented Reality -> Augmented Presence
  - Federated Machine Learning
  - Teleoperation of semi-autonomous vehicles/systems
  - Control of physical systems via *faster-than-real-time* simulation

## **M2C requires a different communication paradigm than E2E**

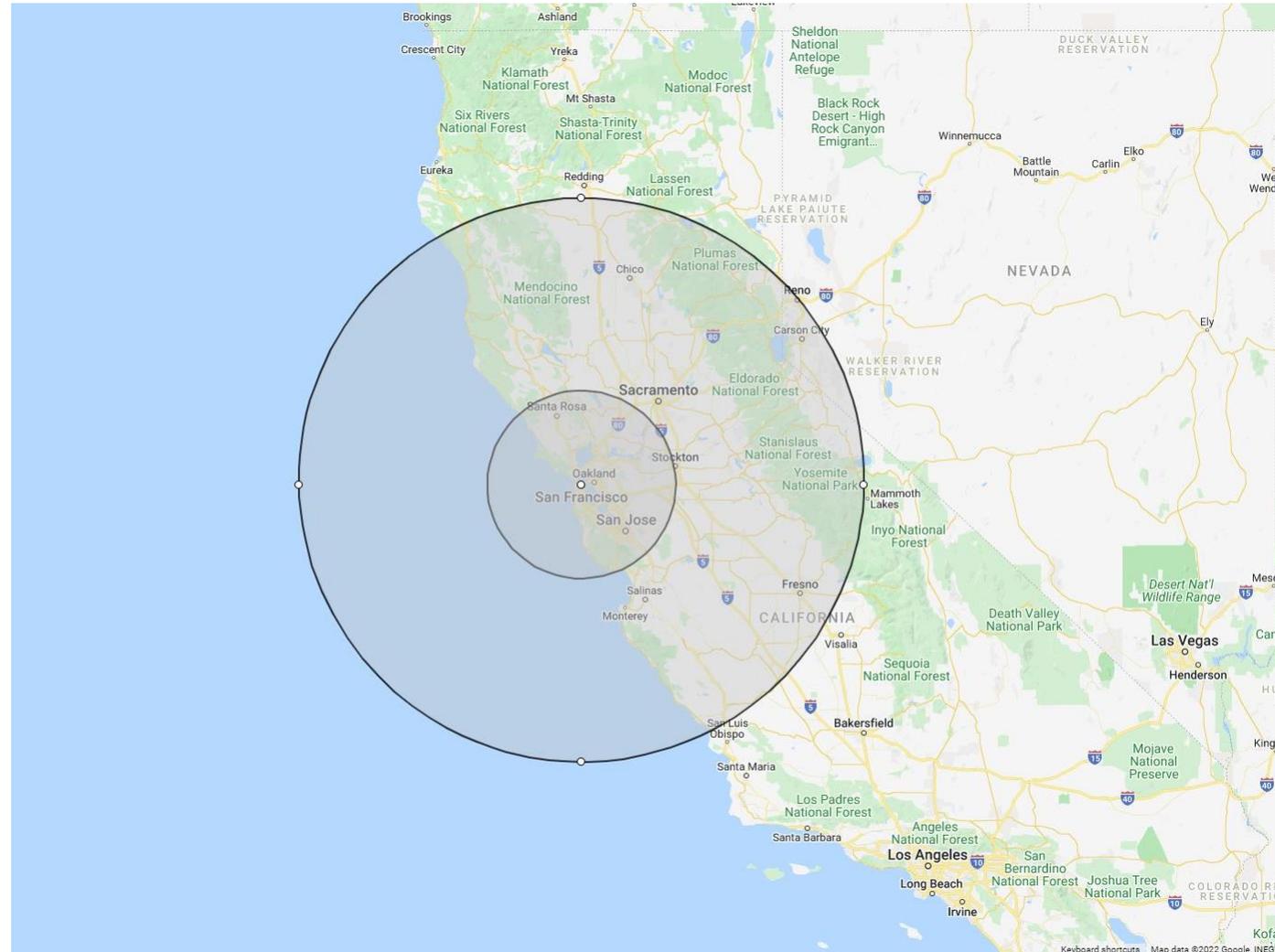
- How will the network facilitate *impedance matching* between the edge and the core

Where Should We Leverage Surprise?



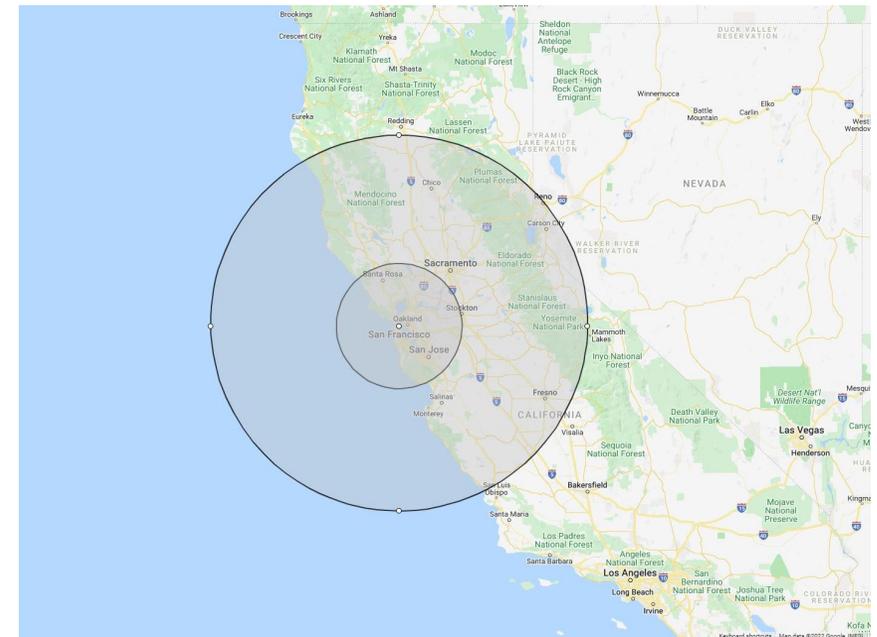
# Near the edge!

$\cong$  within 1 msec of the sensors, actuators & people



# Why Near the Edge?

- Low latency – but only in a realistic and best effort paradigm
  - **Don't even think about Ultra Reliable Low Latency Communications (URLLC)**
  - Probably can't be realized – CAP-like theorem?
  - Wouldn't be relied on if it could



# Why Near the Edge?

- Low latency – but only in a realistic and best effort paradigm
  - Don't even think about Ultra Reliable Low Latency Communications (URLLC)
  - Probably can't be realized – CAP-like theorem?
  - Wouldn't be relied on if it could
- **Surprise: Fuse *new* information from edge devices with *old* information from the cloud**

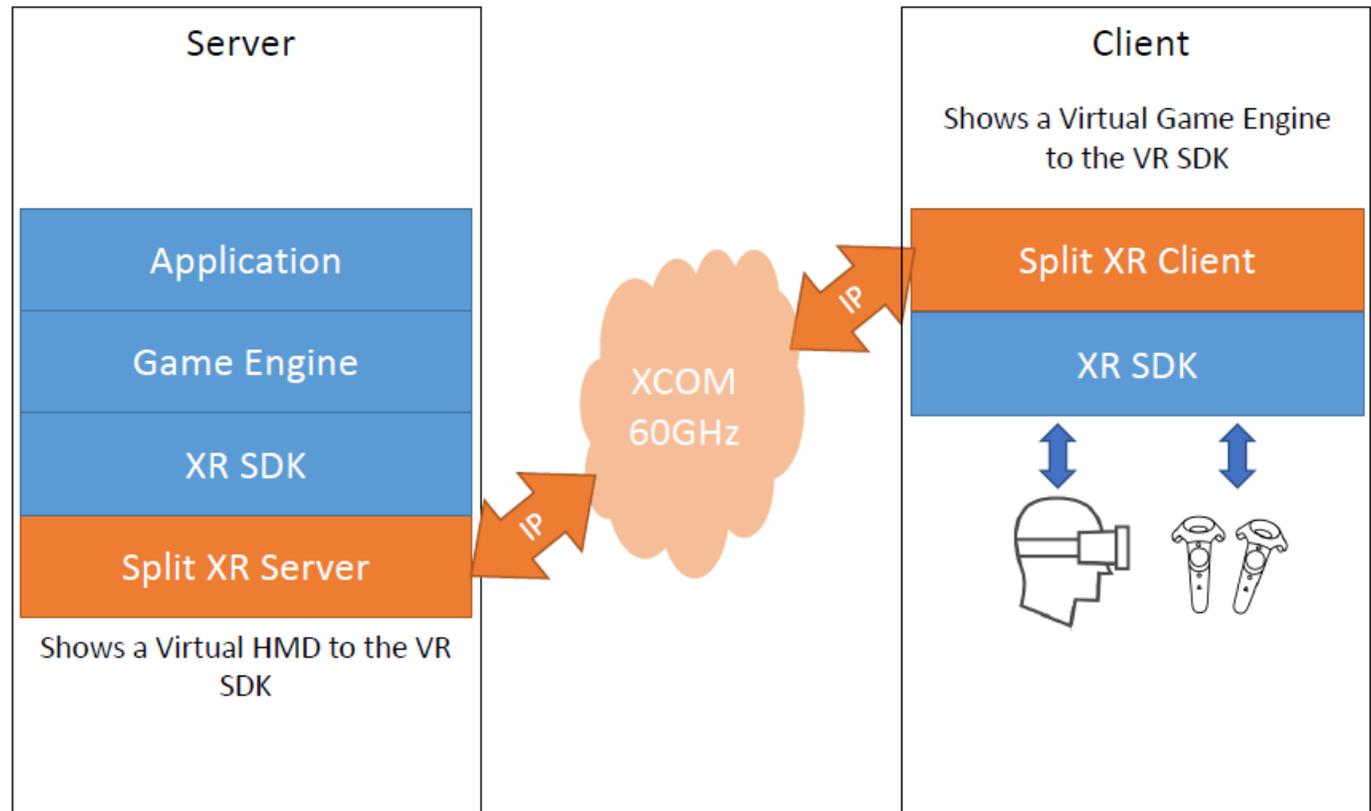
# Example: AR/VR Split Rendering -- Courtesy of XCOM



## Xcalibur: Xcom's Split Rendering for XR

Remoting compute for XR experiences

- Client-Server SW architecture
  - Client requires limited compute to driver HMD
  - Server grows with the requirement of the XR experience
- Transparent to the XR application
  - No need to modify applications
- XR Developer targets the server
  - Free of the resource constrains of mobile HMDs
- All Client-Server data flows are all compressed and error resilient
  - Designed for wireless from the ground up



# IoT and other Real-time Examples

## Smart upstream delivery of new content

- Sensor feeds, anomaly detection, event logs, etc.

## Standing distributed queries that synthesize new / old information

- All of the priors should be pre-digested and staged nearby so that they can be fused with the new data
- Results should automatically be pushed to interested parties

## *Synchronized* multi-cast delivery of *surprise*

- Example: stock trades

Near the Edge is a magic place

Close enough for latency purposes and deep enough for sharing purposes

# Why Near the Edge?

- Low latency – but only in a realistic and best effort paradigm
  - Don't even think about Ultra Reliable Low Latency Communications (URLLC)
  - Probably can't be realized – CAP-like theorem?
  - Wouldn't be relied on if it could
- Surprise: Fuse *new* information from far edge devices with *old information* from the cloud
- **Sovereignty / Privacy**

# Example: Federated Learning



Diagram from Raza, Tran et al, arXiv:2105.12497v1, May 2021

# Why Near the Edge?

- Low latency – but only in a realistic and best effort paradigm
  - Don't even think about Ultra Reliable Low Latency Communications (URLLC)
  - Probably can't be realized – CAP-like theorem?
  - Wouldn't be relied on if it could
- Surprise: Fuse *new* information from far edge devices with *old information* from the cloud
- Sovereignty / Privacy
- **Near the edge is where the cloud providers meet the local communications providers**

Near the Edge is a magic place

# Time for a Re-Partitioning of Concerns?



## Revitalizing the Public Internet By Making it Extensible

*There is now a significant and growing functional gap between the public Internet... and a new generation of more sophisticated private networks.*

*...we argue for the creation of an Extensible Internet (EI) that supports in-network services that go beyond best-effort packet delivery.*



**Scott Shenker**

2021. Revitalizing the public internet by making it extensible. *SIGCOMM Comput. Commun. Rev.* 51, 2 (April 2021), 18–24. DOI:<https://doi.org/10.1145/3464994.3464998>

## From Cloud Computing to Sky Computing

*We consider the future of cloud computing and ask how we might guide it towards a more coherent service we call sky computing.*



**Ion Stoica**

2021. From cloud computing to sky computing. In *Proceedings of the Workshop on Hot Topics in Operating Systems (HotOS '21)*. Association for Computing Machinery, New York, NY, USA, 26–32. DOI:<https://doi.org/10.1145/3458336.3465301>

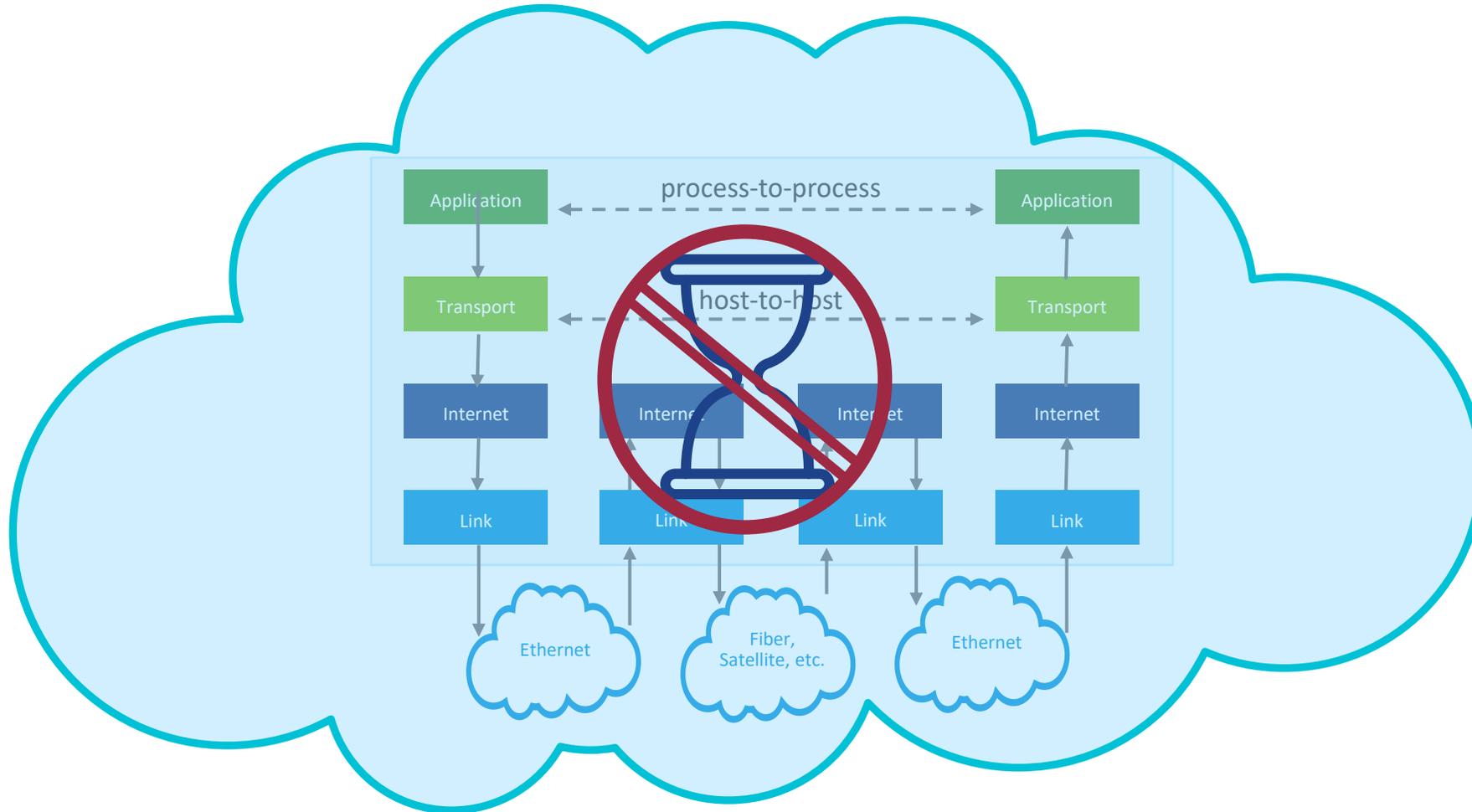
What's the Programming Model?



# Think Beyond the *Narrow Waist*



Bob



Alice

## Tools of our Trade

Computation in the network

Storage in the network

Information Theory / Surprise

# Summary: Implications for NextG

Re-imagine communication and computation through the lens of *surprise*

- Fusion of *new* information from the edge with *old* information from the cloud
- This likely requires a new programming paradigm

Enable computation and storage “near the edge”

- Get really good at hosting application layer relays “near the edge”
- Lose the “narrow waist” obsession

Move beyond the 3GPP mindset – we need a repartitioning of interests

- Stop the madness wrt URLLC
- Focus on the place where the hyper-scalers meet the communication service providers

# Let's face reality...



AND enable the next generation of network rebels 😊

