Introduction to Science Networking

Computational HEP Traineeship Summer School 2024

May 20th, 2024

Asif Shah

Software Developer II

(U.S. CMS Networking R&D)

Email: sshah@fnal.gov

Outline

- History of Internet & HEP Networking
- Building Blocks Network Technology
- Scientific Collaboration and High-Speed Networks
- Science Networking R&D
- Q & A Session

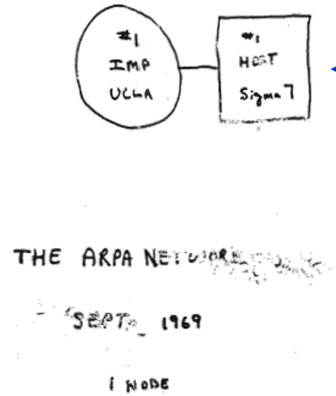
History of Internet & HEP Networking

The Beginning…

- (1969) US Advanced Research Projects Agency (ARPA) funds development of new communications network:
  - Not circuit-oriented
  - Data moved in independent packets called datagrams
  - Store-and-forward
  - Hop-by-hop dynamic routing
- Project was a collaboration between US research universities, US high-tech industry, & US military
- Primary objective = highly efficient network
  - Background objective = flexible, adaptable network (potential military use…)

ARPAnet (1969-1970s):

- Used mini-computer (IMP) as router for datagrams (packets)
- Communications protocol - NCP
- End system (Host) services:
  - Initially - remote logon
  - Later - file transfer & email

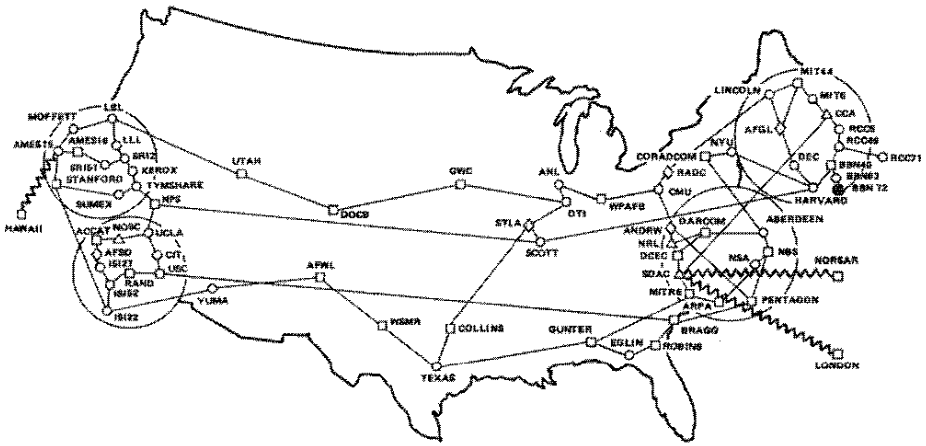


(design)

[UCLA](#): 1st ARPAnet Node

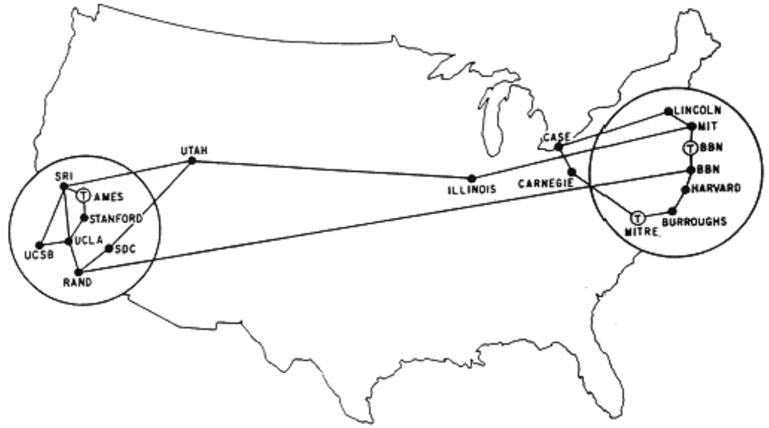
(IMP)

Mini-computer?



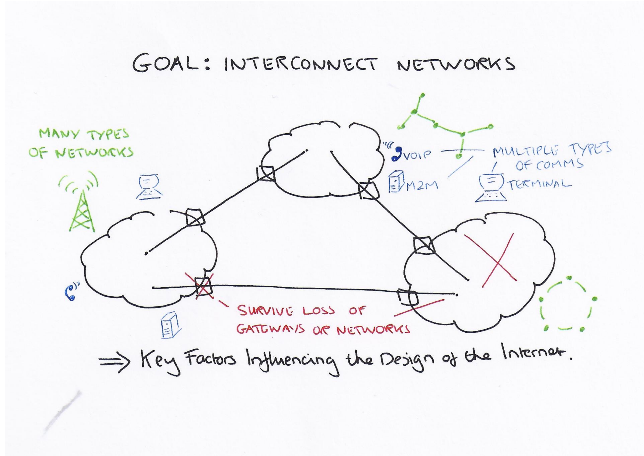
[ARPAnet](#): Sept. 1971 ---→

←--- ~1980



DARPA Internet (late1970s through 1980s):

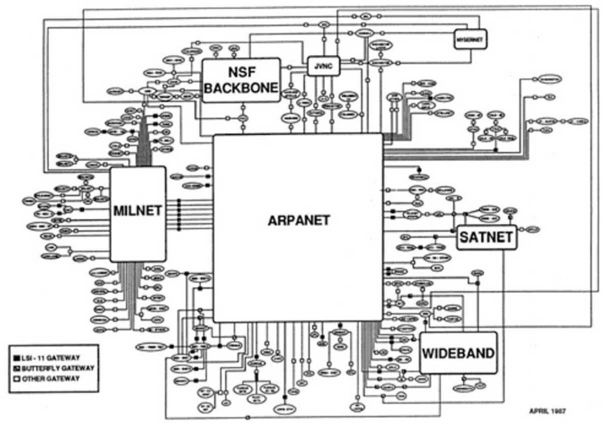
- DARPA (successor to ARPA) directed focus on interconnected networks, starting in 1973:
  - **Separate administrative domains**
  - Underlying communications technology independence/flexibility
- Funded development of a new suite of communications protocols:
  - TCP – reliable connection-oriented host-to-host delivery
  - IP – connection-less, not-guaranteed handling by the network
  - Jan 1983 – TCP/IP replaces NCP



(1973)

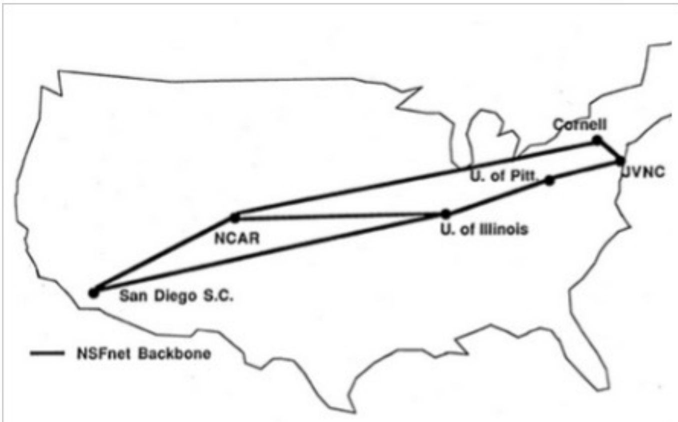
leads to

(1987)



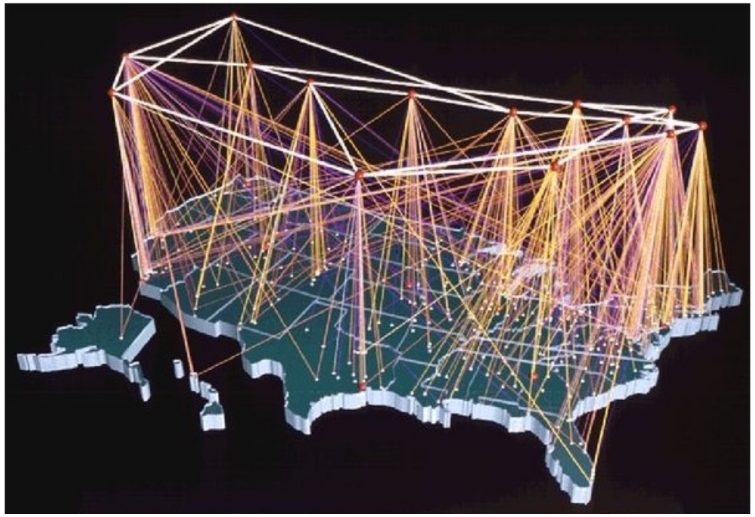
NSF Takes The Internet into Mainstream (1980s):

- NSF creates NSFnet, to connect its five supercomputer centers:
  - Objective is to support research, not be a research project itself
  - Complaint: “It’s analogous to a superhighway with no on/off ramps…
- NSF expands NSFnet’s scope:
  - Added backbone sites, which were in more strategic locations
  - Seeded regional networks to connect universities to NSFnet:
    - Regional <-> regional traffic quickly dominates



(1986)

leads to



(1993)

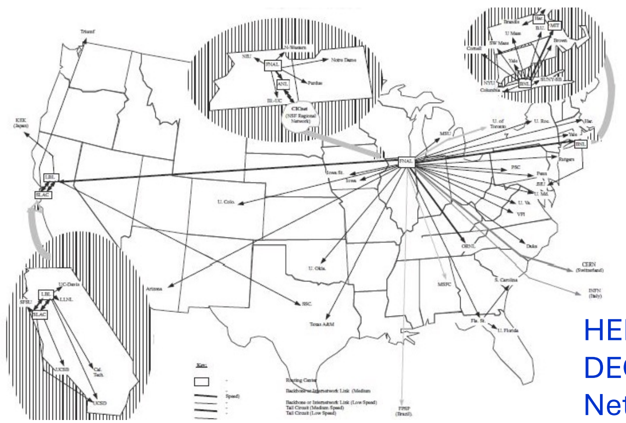
The Internet Goes Global, then Commercial:
(early-mid1990s)

- In Europe:
  - Telecom-aligned national entities pushed for International Standards Organization-developed network protocols, not TCP/IP:
    - Concerns over US dominance in development of TCP/IP standards
  - Multi-protocol inter-European backbone (EARN) established:
    - Supported ISO Open Systems Interconnect and TCP/IP
    - TCP/IP wins out, based on demand…
  - European national IP networks connect with NSFnet to form core of global internet
- US Government "privatizes" the NSFnet backbone (1994):
  - Four commercial service providers
  - Lays groundwork for evolution of commercial side of the internet

Meanwhile, HEP developed its own network:
(mid-late1980s)

- 1983-ish - Remote terminal network:
  - University researchers use remote (dumb) terminals to access to HEP computers
  - Asynchronous data switches at experiment sites connected to form network:
    - Unofficially dubbed **HEPnet**
- 1984-1990 – DEC VAX mini-computers proliferate:
  - Proprietary DECnet protocol replaces async. terminal communications
  - HEP DECnet grows to a pre-internet global network (~10k nodes)

Remote Terminal Network (1983)



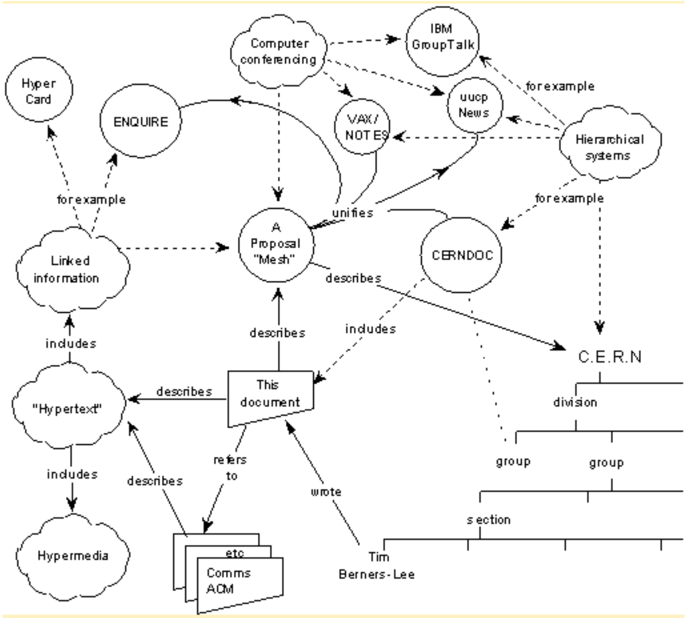
HEPnet DECnet Network (~1990)

US Dept. of Energy consolidates its networks:
(1990-ish)

- Mid-1980s – DOE is supporting two science networks:
  - HEP DECnet network (using vendor proprietary protocols)
  - Magnetic Fusion Energy (MFE) network (using custom in-house protocols)
  - Other DOE science disciplines on the horizon will need a network too
- 1985 – DOE decides to unify its current & future network support:
  - A single shared network facility
  - Interoperable, based on TCP/IP protocol suite
- 1987 – ESnet (Energy Sciences network) is born:
  - HEP DECnet transitions to ESnet as a network backbone
    - 1989 - ESnet converts to multiprotocol routers (TCP/IP & DECnet)
  - MFEnet converts its in-house protocol to be TCP/IP-based

HEP has contributed to Internet evolution as well

- 1990 – CERN's Tim Berners-Lee designed a data sharing framework for the Internet
- Three core elements:
  1. Hypertext markup language (HTML)
  2. Uniform Resource Identifier (URI)
  3. HyperText Transfer Protocol (HTTP)
- Collectively creating The World-Wide Web
- It's been described as the Internet's "killer application"
  - CERN made code available royalty-free
  - Sparked Internet-wide wave of innovation and collaboration



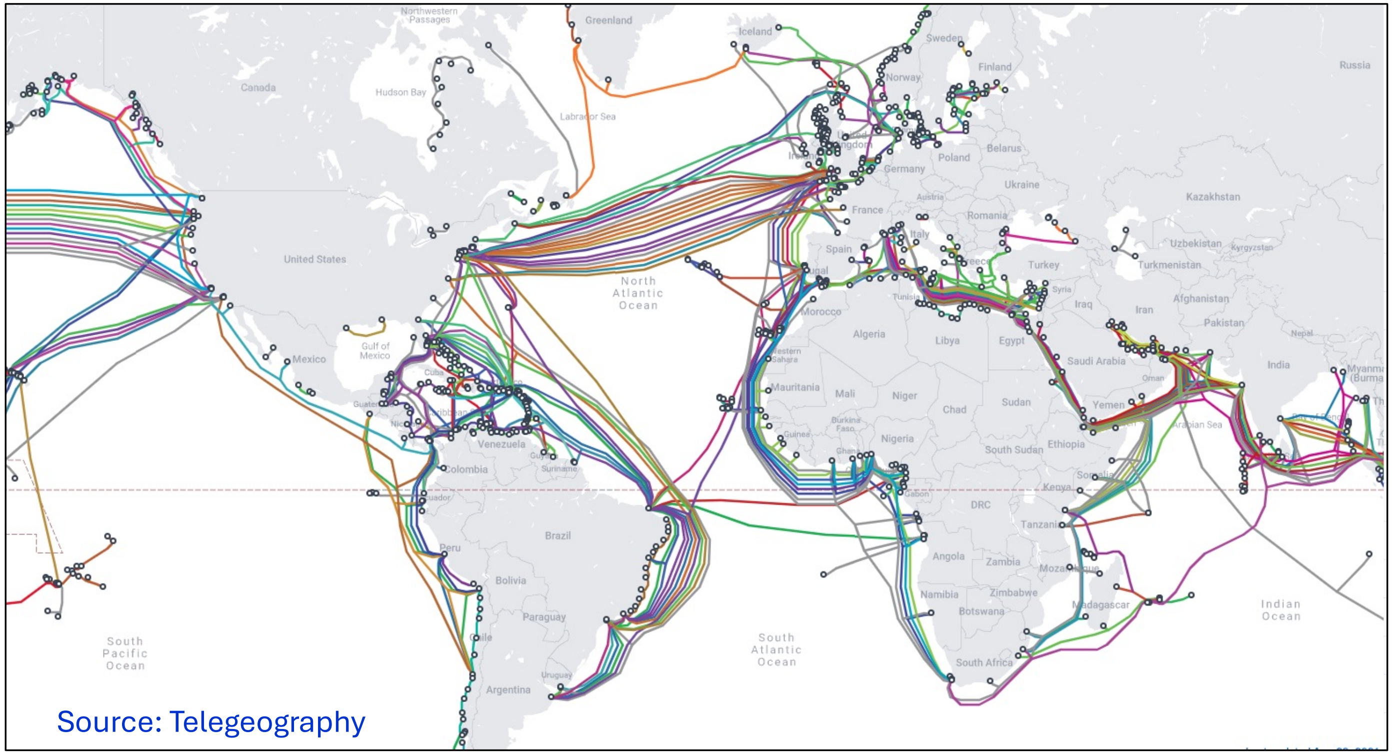
Schematic Diagram from Berners-Lee's Original Proposal

Building Blocks of Networking Technology

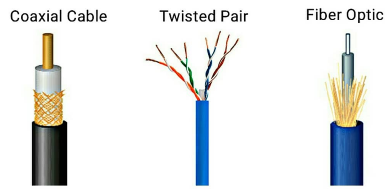
Building Blocks of Networking Technology (1)

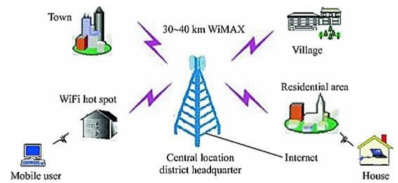
1. Transmission Media – Wired & wireless media
   - Features: Speed, Cost, Distance, & Secure

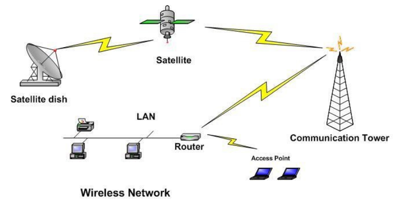
Internet Cables Undersea Worldwide Map – Backbone!



Source: Telegeography



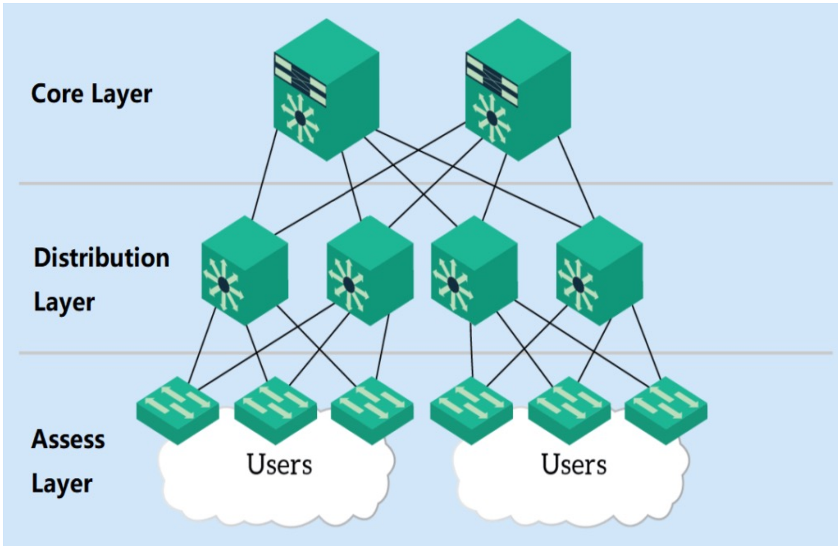




Building Blocks of Networking Technology (2)

2. Larger scale networks – Number of various networking **DEVICES** involved.
   - NICs, Modems, Switches, Repeaters, Routers, Gateways, Firewall, etc.

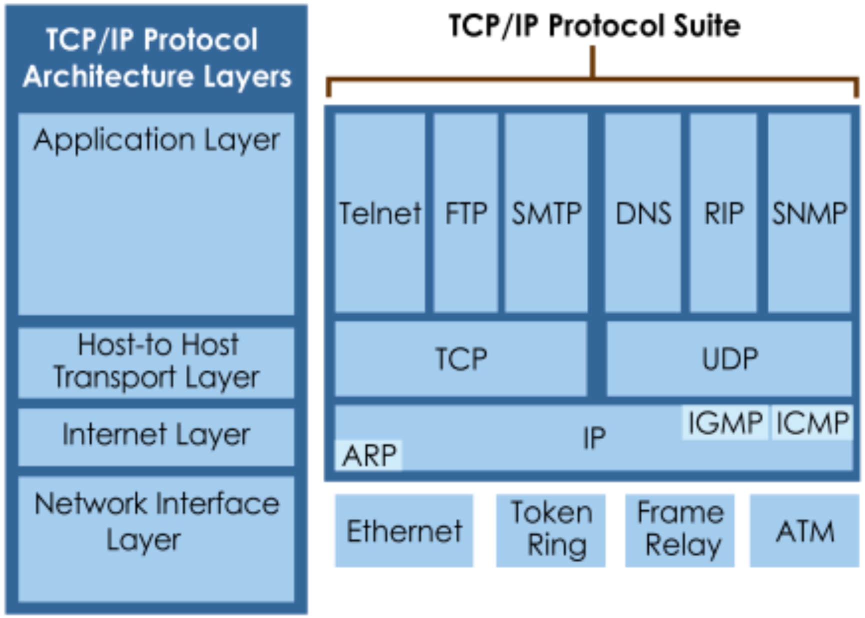
Hierarchical Network: Core, Distribution, Access Layers

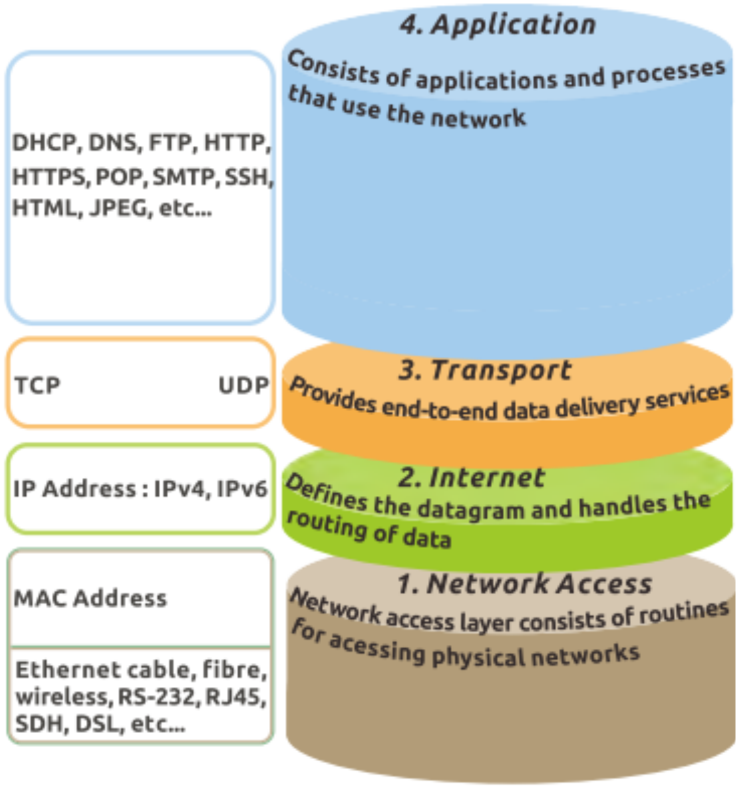


Building Blocks of Networking Technology (3)

3. How to communicate in networking environment?

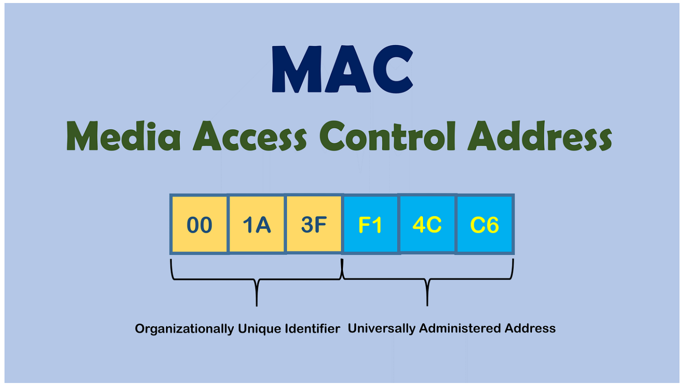
- TCP/IP protocol suite

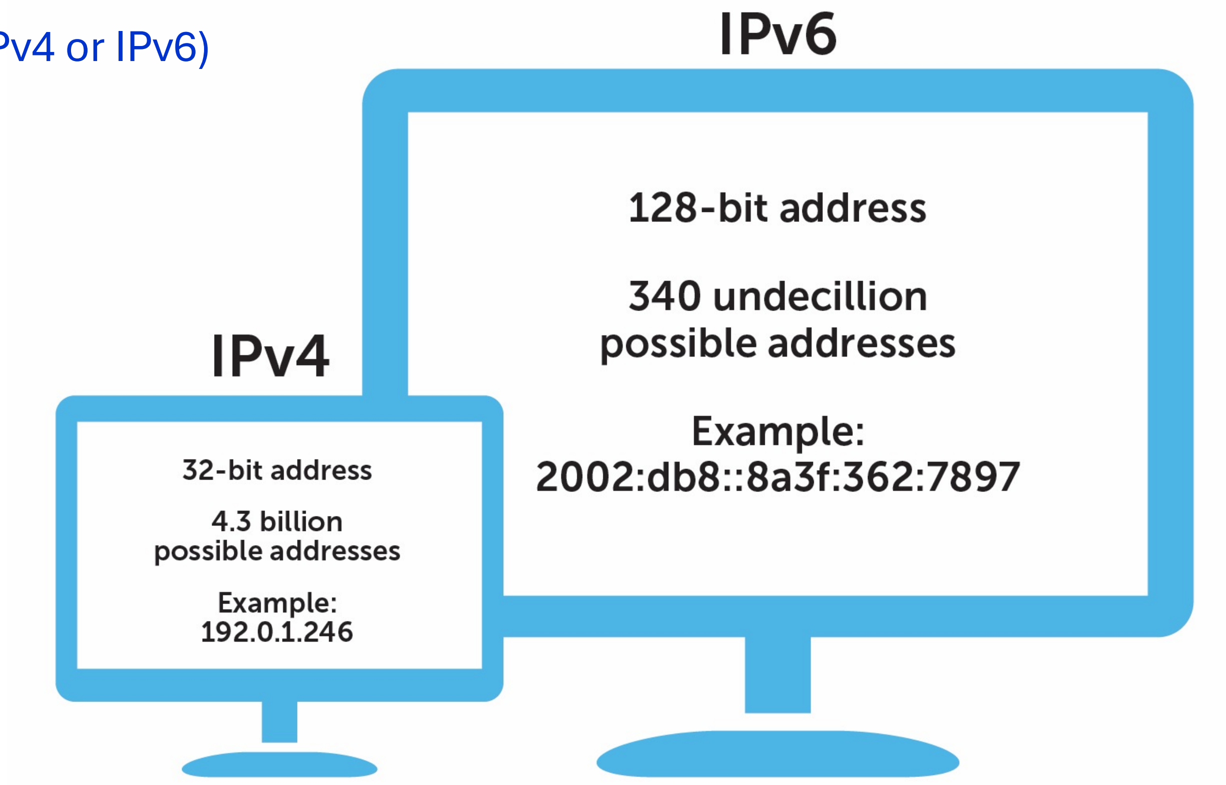




Building Blocks of Networking Technology (4)

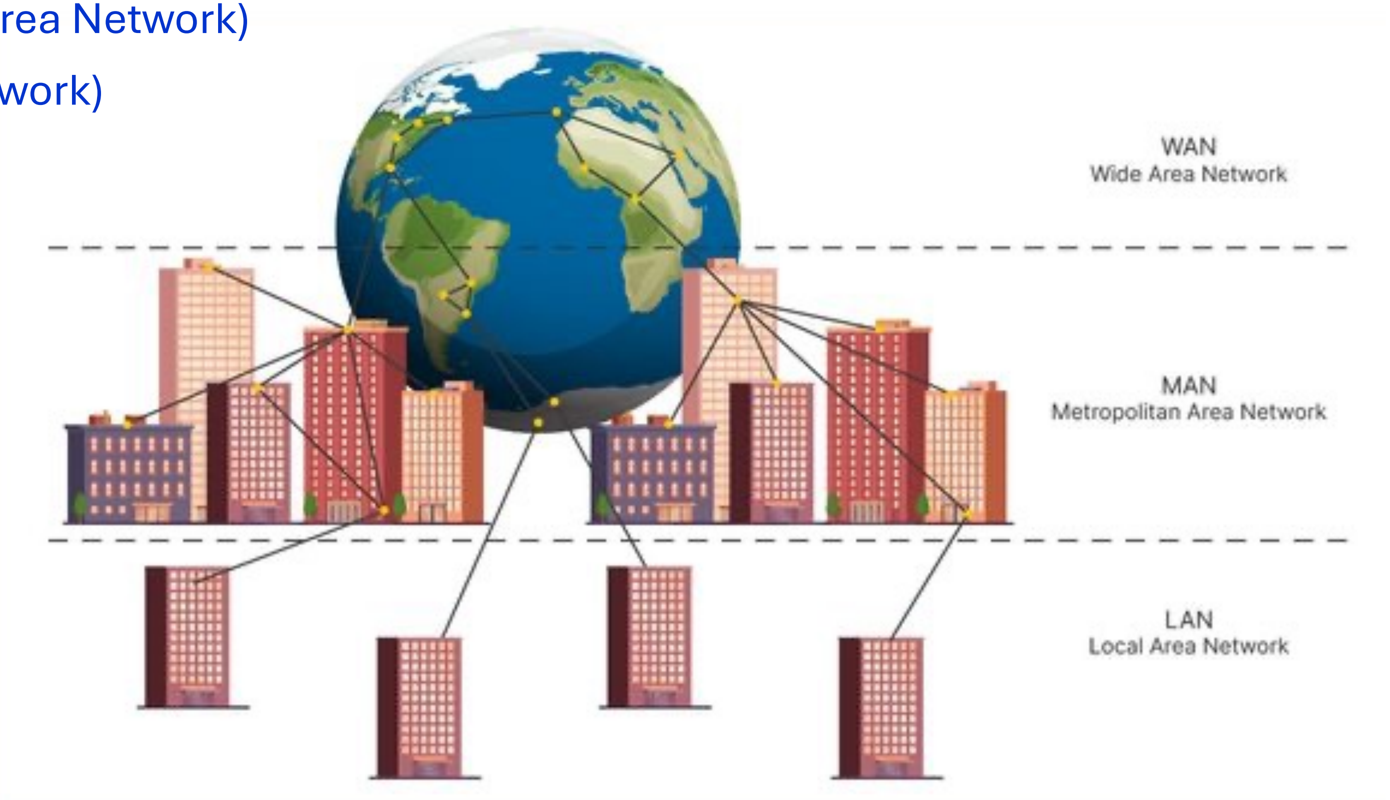
4. Identification of computers/devices and networks?
   - Physical Address – MAC address
   - Logical Address – IP Address (IPv4 or IPv6)





Building Blocks of Networking Technology (5)

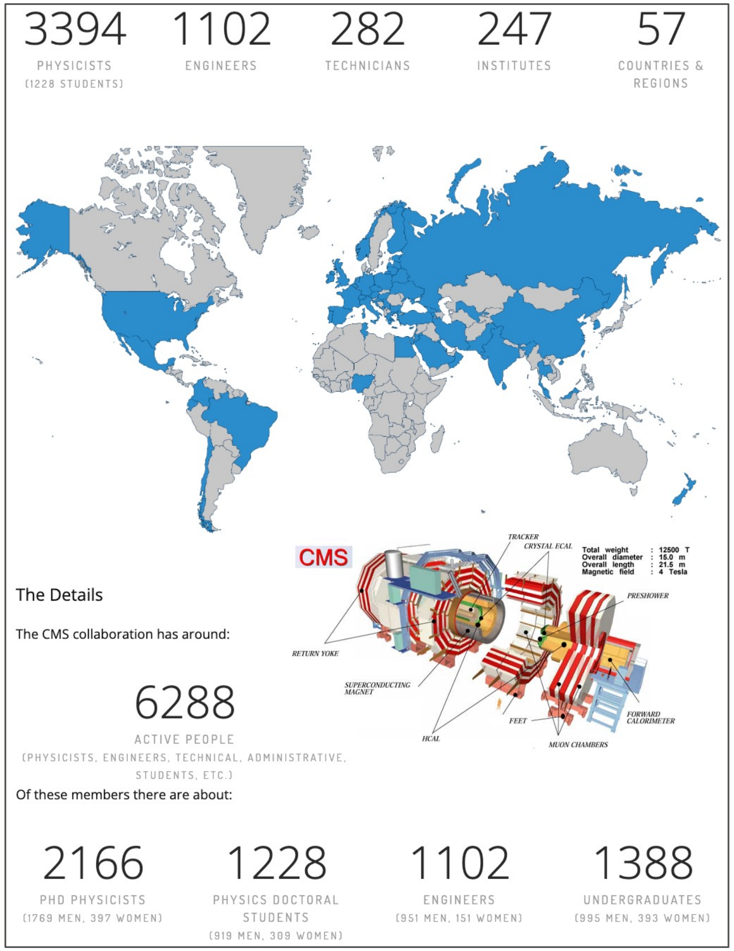
5. Geographic distribution of networks?
    - LAN (Local Area Network)
    - MAN (Metropolitan Area Network)
    - WAN (Wide Area Network)



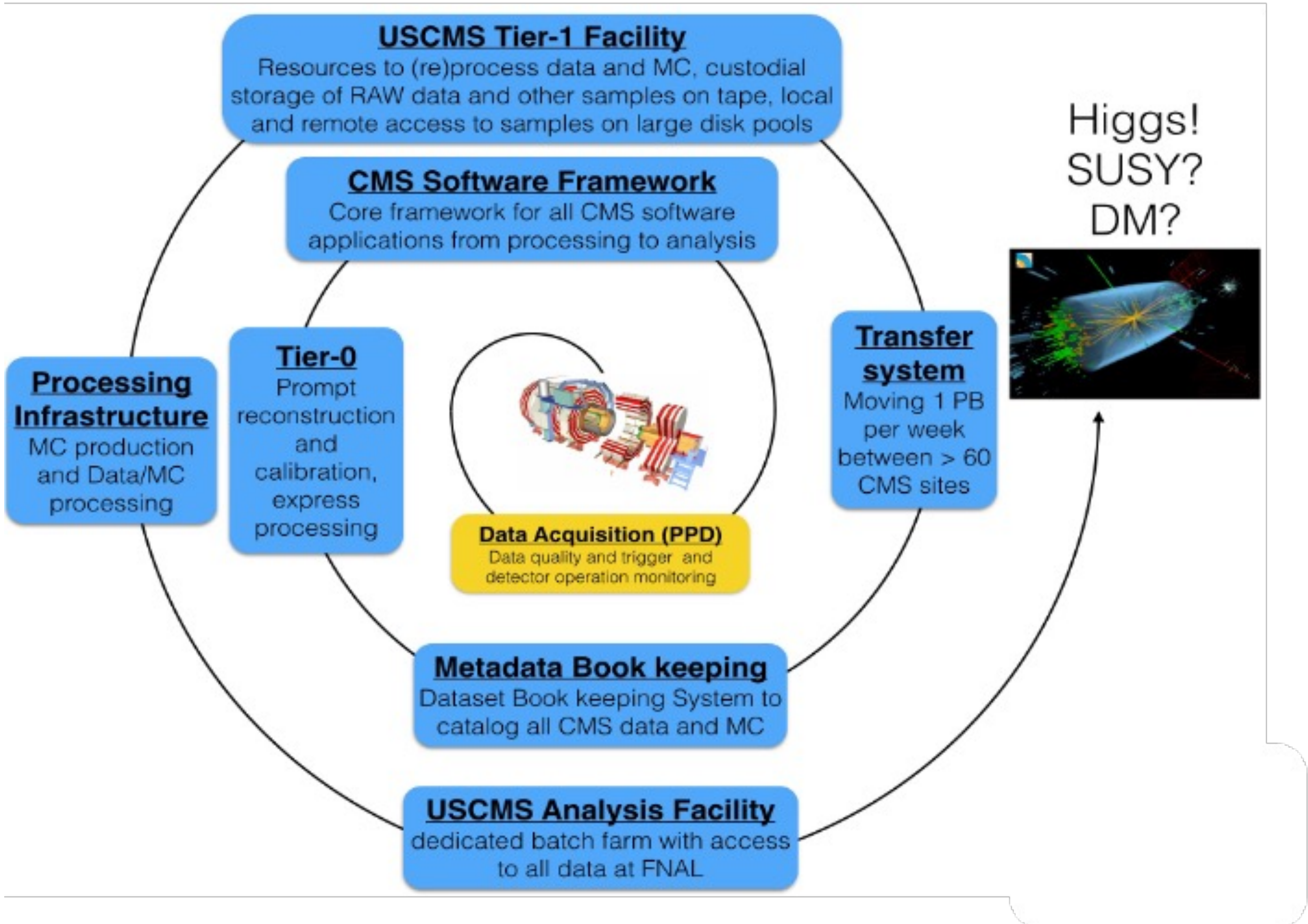
Scientific Collaboration and High-Speed Networks

Scientific Collaboration – Example (CMS Experiment)

- Extreme data volumes & velocities
- Large collaborations of global scale
- Highly distributed computing environment(s): Federated
- Over 200 Institutes from 57 countries
- Over 3300 physicists
- Worldwide 7 Tier-1 sites and more than 50 Tier-2 sites
- CMS compute needs are mainly covered by WLCG resources, a global collaboration of about 170 computing centers, aggregating 1M CPU cores and 1 EB of storage (disk and tape)

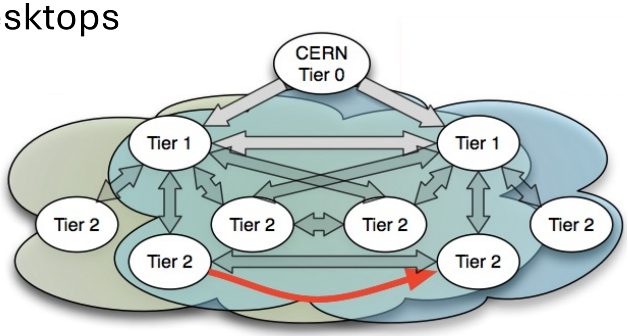


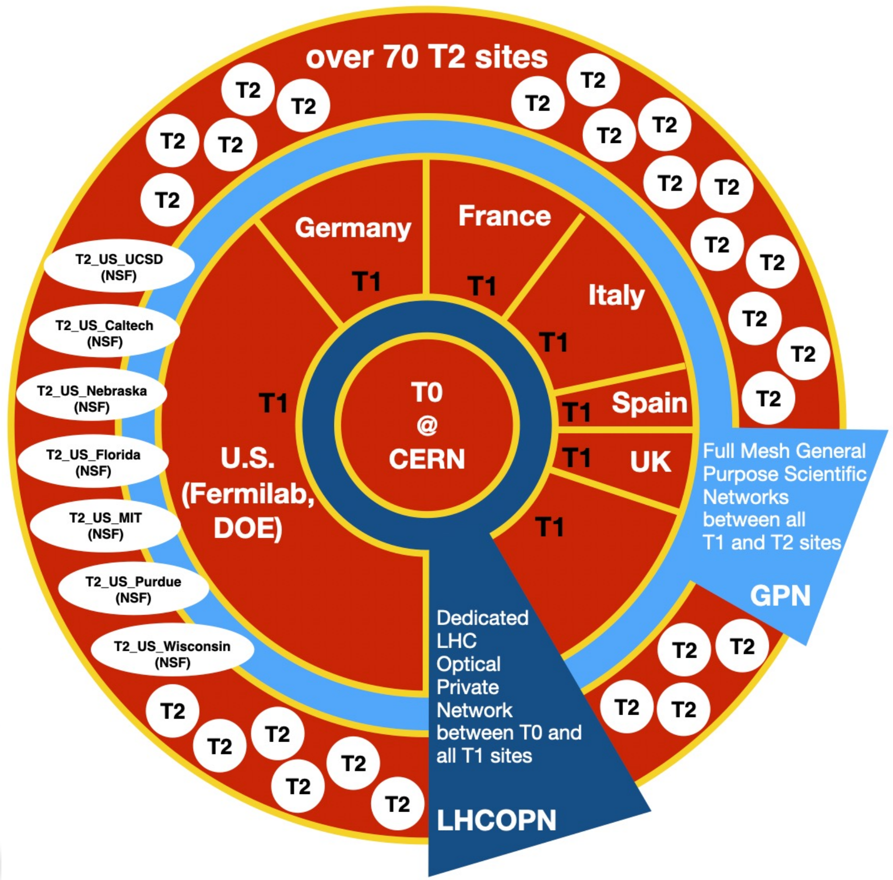
Scientific Collaboration – CMS Computing Lifecycle



Scientific Collaboration – Tier Structure & Networking

- Tier structure for computing:
  - Tier 0 = CERN
  - Tier 1 = National data centers for event reconstruction & archiving
    - **Connected using LHCOPN Networks**
  - Tier 2 = Computing facilities for Monte Carlo production & event analysis
    - **Connected using LHCONE/Other R&E Networks**
  - Tier 3 = Collaboration sites
    - **Connected using Regional R&E Networks**
  - Tier 4 = Physicist desktops

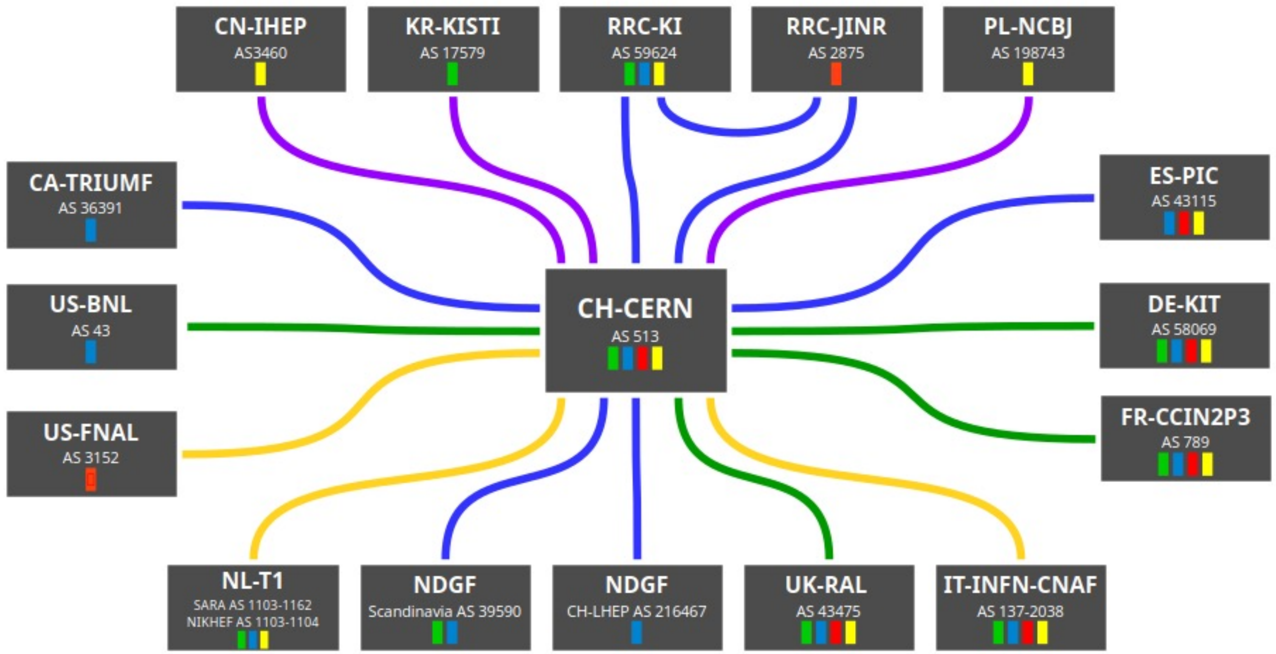


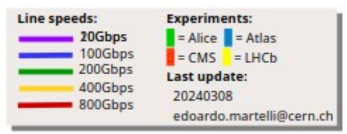


High Speed Scientific Networks – LHCOPN (LHC Optical Private Network)

LHCOPN

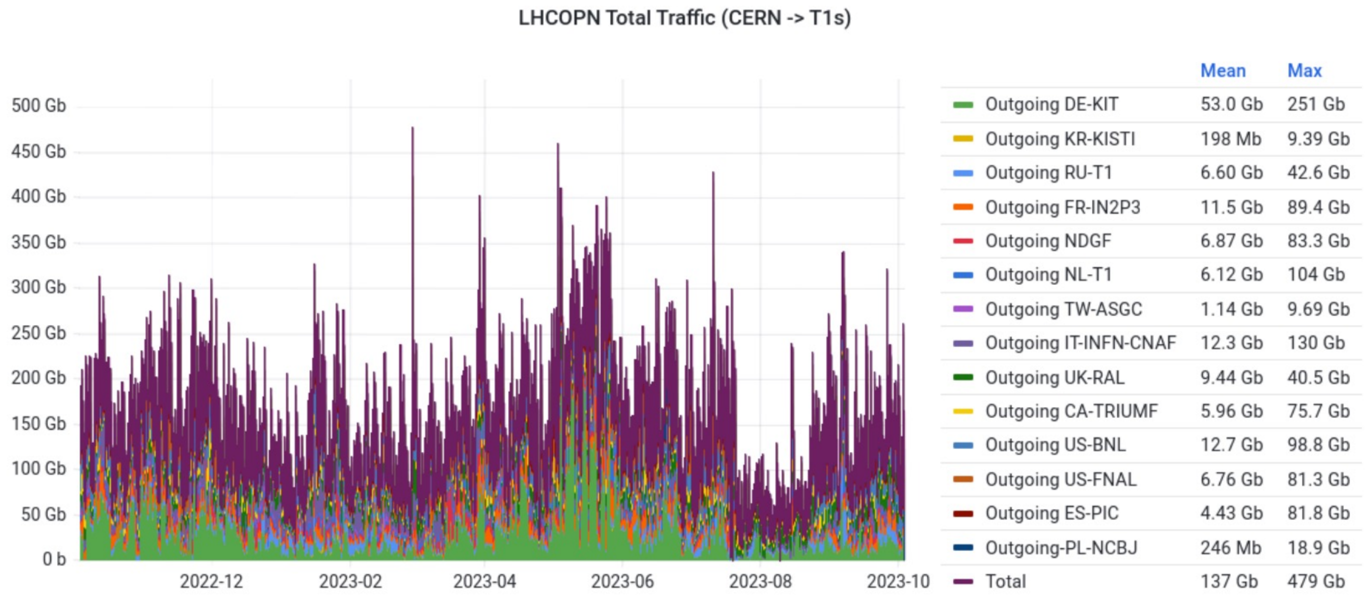
- The LHC Optical Private Network (LHCOPN), linking CERN and the Tier 1s:
  - 18 sites for 17 Tier1s + 1 Tier0
  - PL-NCBJ just joined, CN-IHEP and NDFG-LHEP in the process to connect
  - 15 countries in 3 continents
  - Dual Stack IPv4–IPv6
  - 2.1 Tbps to the Tier0



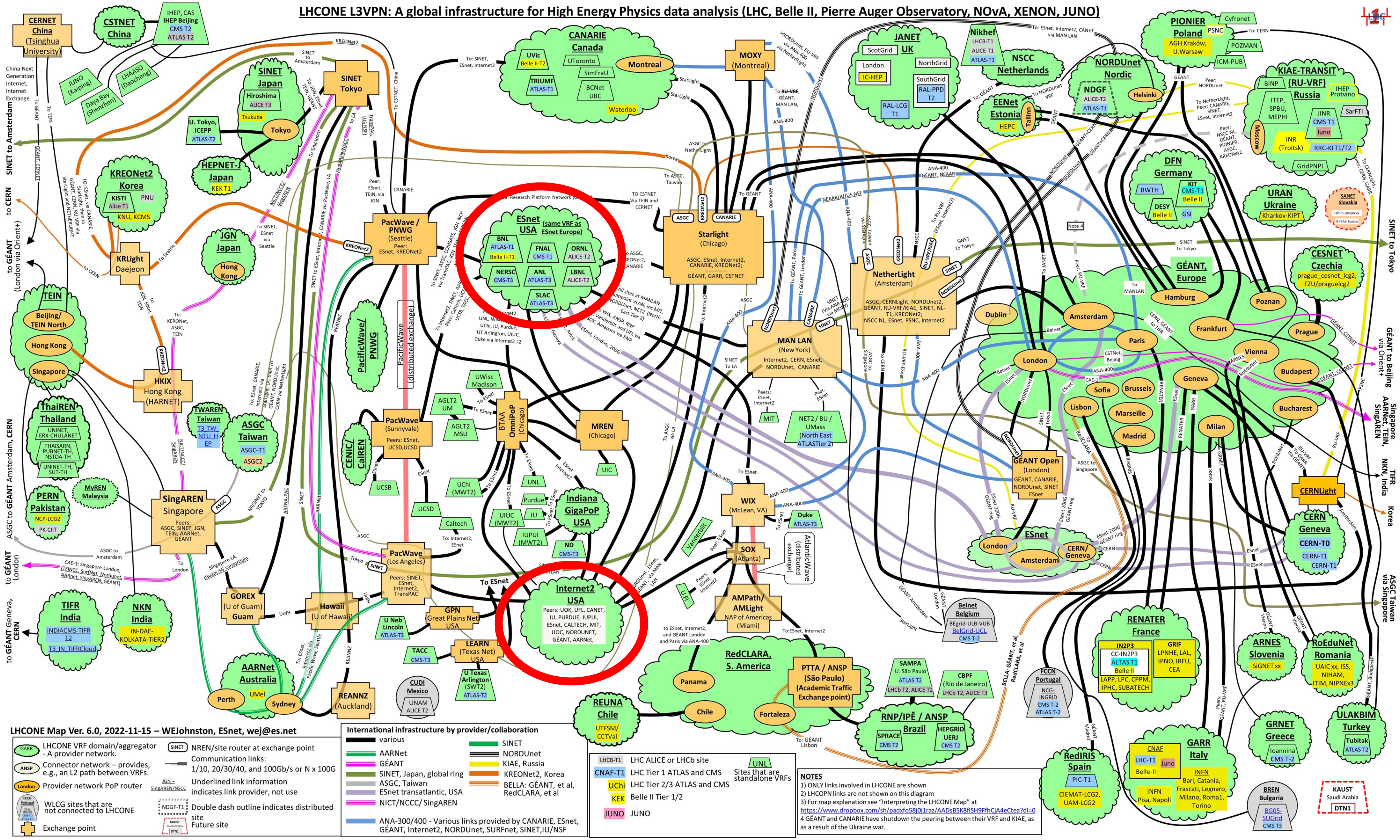


High Speed Scientific Networks – LHCOPN (LHC Optical Private Network)

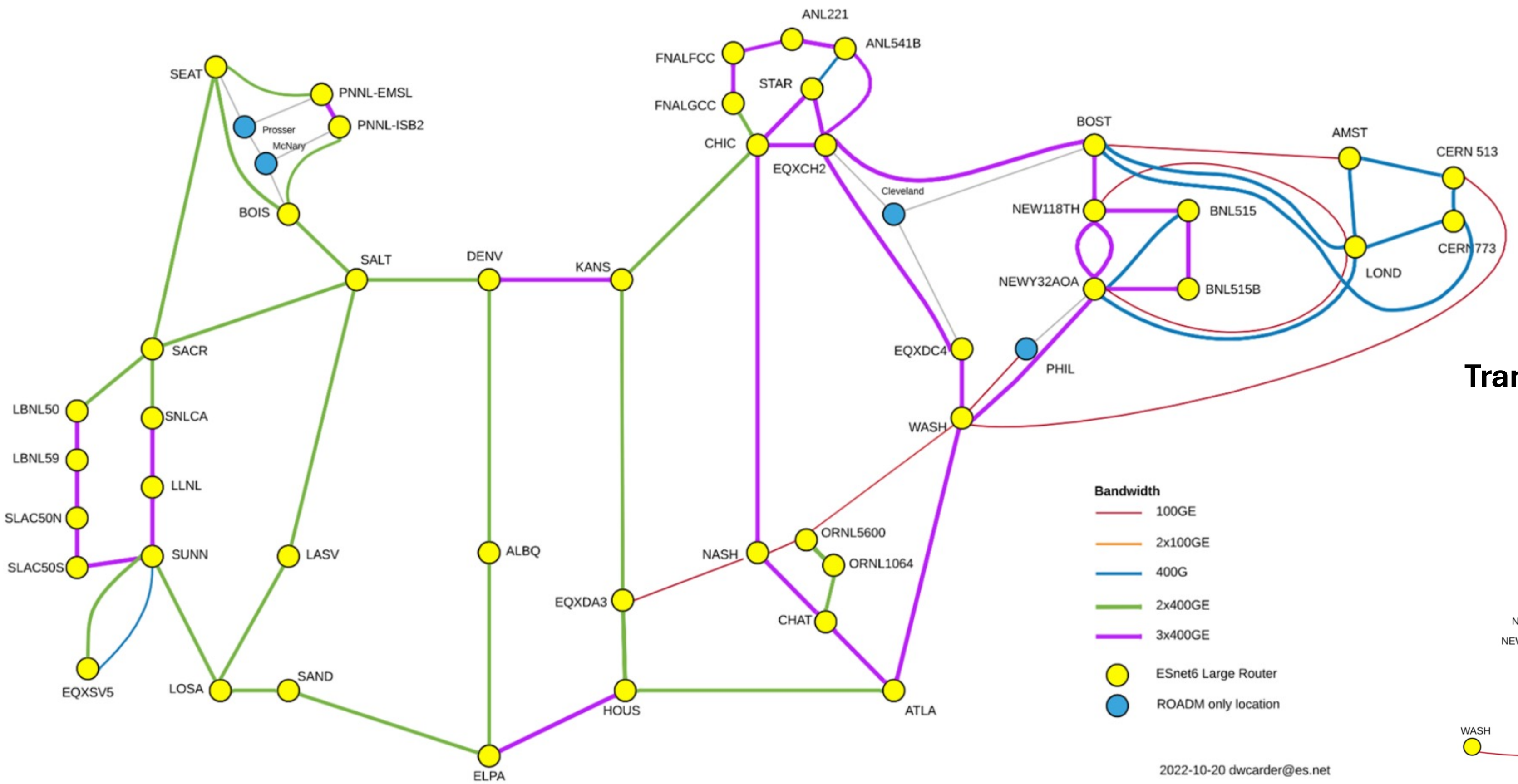
- LHCOPN Traffic – last 12 months
  - Moved ~540 PB in the last 12 months
  - +18% compared to previous year (457PB)
  - Peak at ~479Gbps



High Speed Scientific Networks – LHCONE (LHC Open Network Environment)

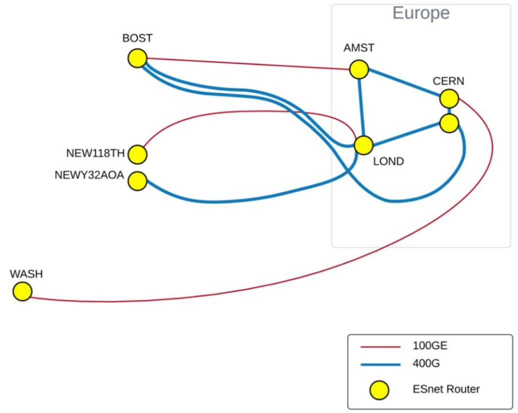


ESnet Backbone Network – Q3 2023



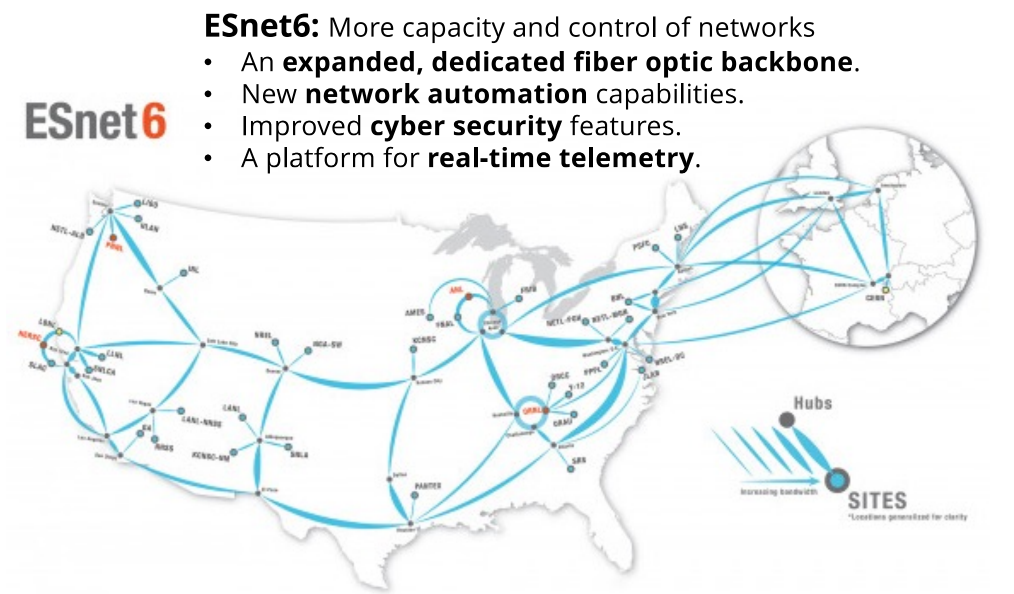
Trans-Atlantic links capacity:

- 1.5Tbps in 2024 (Achieved)
- 3.2Tbps in 2027 (Future)



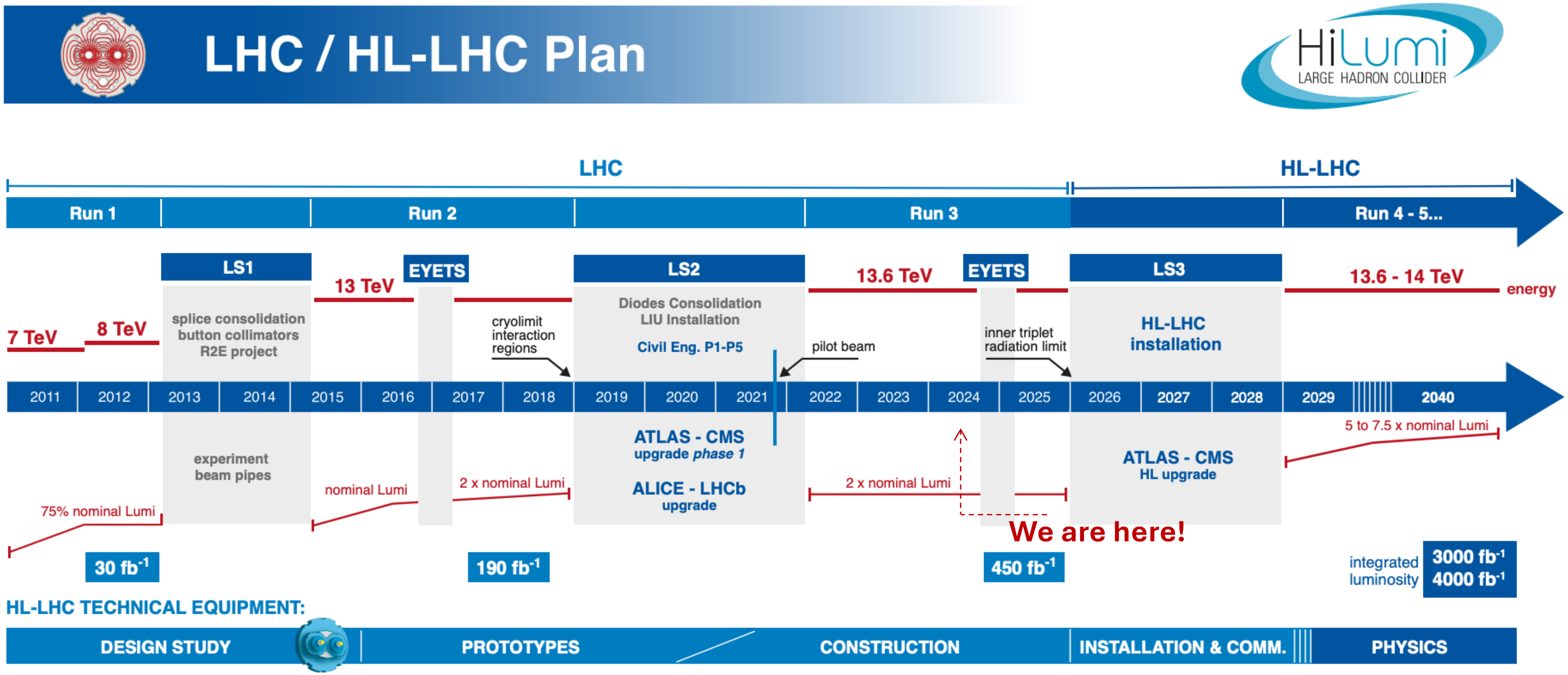
ESnet is now a Globally-Preeminent R&E Network

- ESnet now in its 6th version
  - ESnet manages its own optical fiber system in US
    - ‘n’ x 400GE links between routers
  - Extensive trans-Atlantic connections to Europe
    - Including trans-Atlantic spectrum
  - DOE supercomputing center connectivity:
    - ALCF, OLCF, NERSC
  - Rich peering with other networks:
    - Including cloud providers
- HEP is ESnet’s biggest customer:
  - ~65% of ESnet traffic is HEP

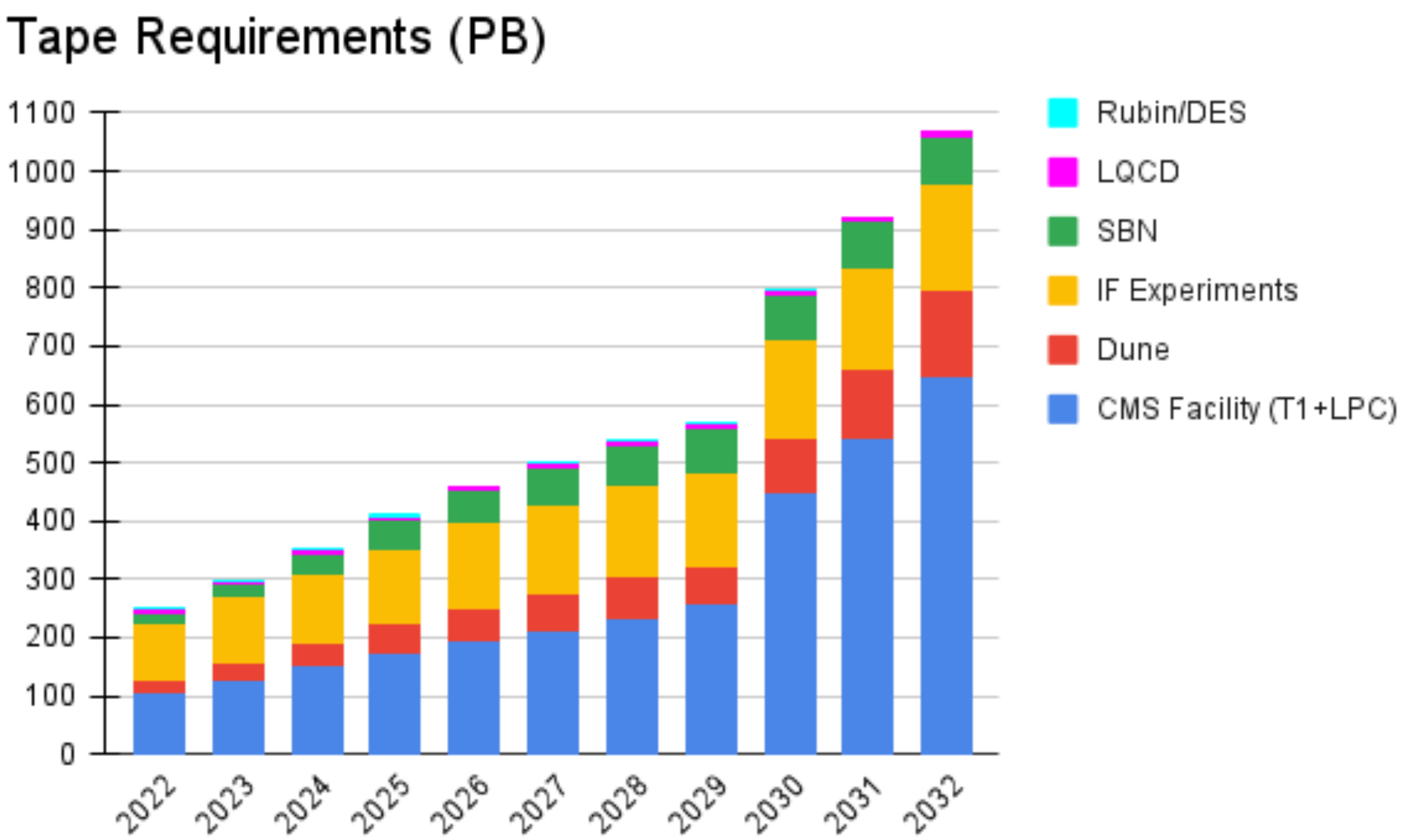


Science Networking R&D

Science Networking R&D – LHC/HL-LHC Schedule

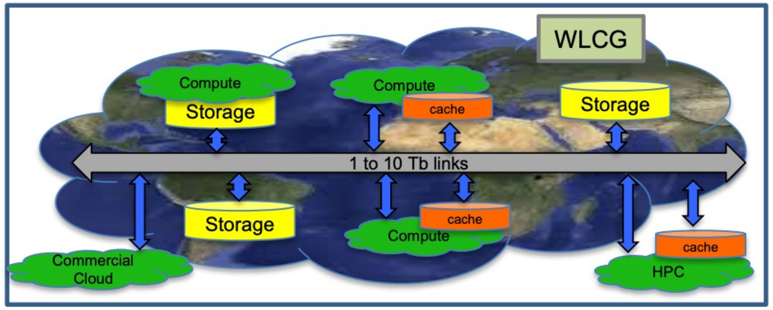


Science Networking R&D – Projected Data Volume on Tape



Science Networking R&D – Emerging Science Networking Research

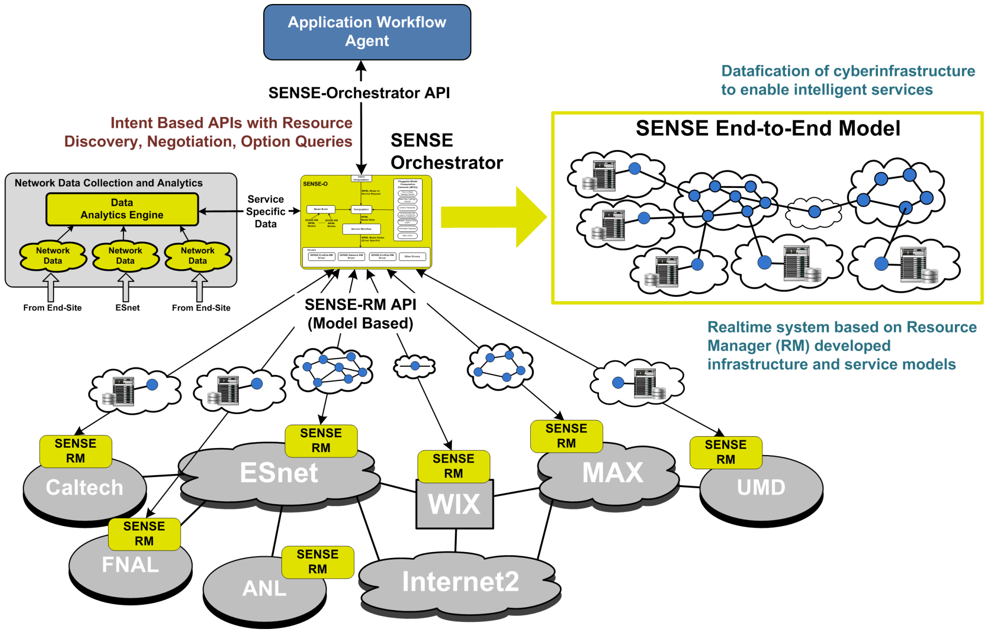
- It is expected that the exabytes of new data during the Hl-LHC era (2029-2040)
- Networks are the glue for distributed computing, connecting processing with storage
- Relies on ~Tb/s networks capacity at backbone and R&E networks
- But it is important to demonstrate the ability to use the full capabilities of the network at all level:
  - LHCOPN/LHCONE/R&E Networks/Campus networks/etc.



- Emerging Networking R&D:
  1. Network management – Software-Defined Networks & Network Orchestration
  2. Network visibility – SciTags (Packet Marking & Flow Labeling)
  3. Network usage optimization – Packet Pacing & Traffic Shaping
  4. Network Monitoring – Health monitoring
  5. AI/ML for Networking – Traffic analysis & predictions

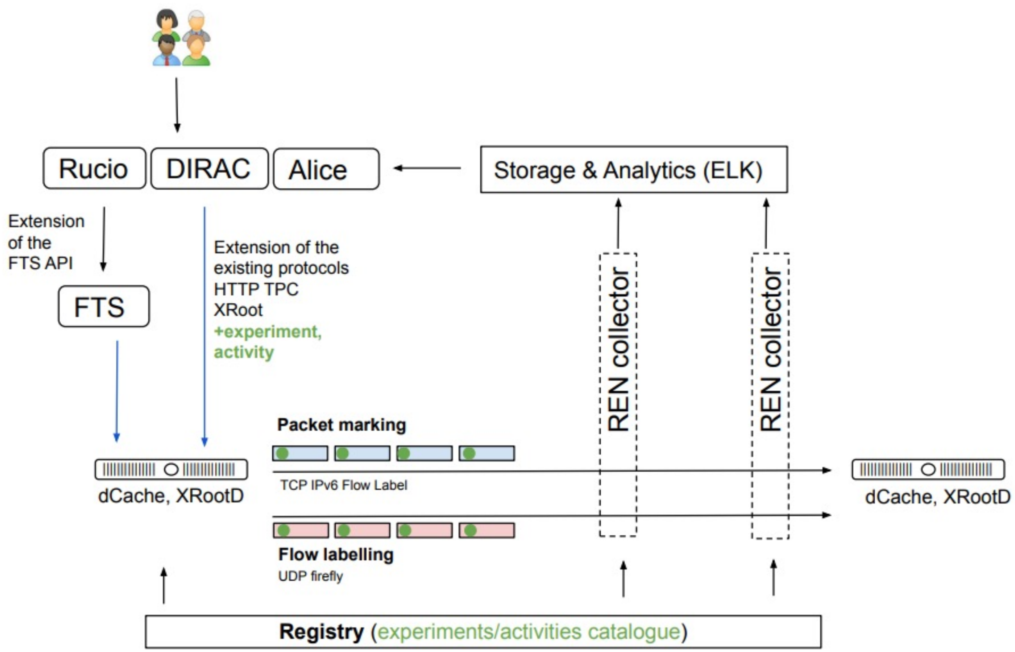
Emerging Networking R&D – Network Management

- A challenge for HEP storage endpoints is to utilize the network efficiently and fully.
- Software-Defined Network (SDN) & Network Orchestration
    - Virtualized overlay network paths (or even networks)
    - Intended to provide specific service needs (ie., QoS) and/or isolation
    - GNA-G (Global Network Advancement Group)
    - The SENSE ESnet project is serving as a reference implementation, FNAL is part of testbed
- The NOTED project (CERN) is also an example of a practical way to effectively utilize available paths to better distribute network load. FNAL also part of testbed



Emerging Networking R&D – Network Visibility (SciTags)

- Scientific Network Tags (Scitags) is an initiative promoting identification of the science domains and their high-level activities at the network level.
- Provide standardized means of information exchange on network flows between experiments, sites and network providers.
- Improve experiments and sites visibility into how network flows perform within network segments.
- Get insights into how experiments are using the networks and benefit from additional data from the network providers.



Emerging Networking R&D – Network usage Optimization

- Pacing/Shaping WAN data flows to optimize the overall network traffic.

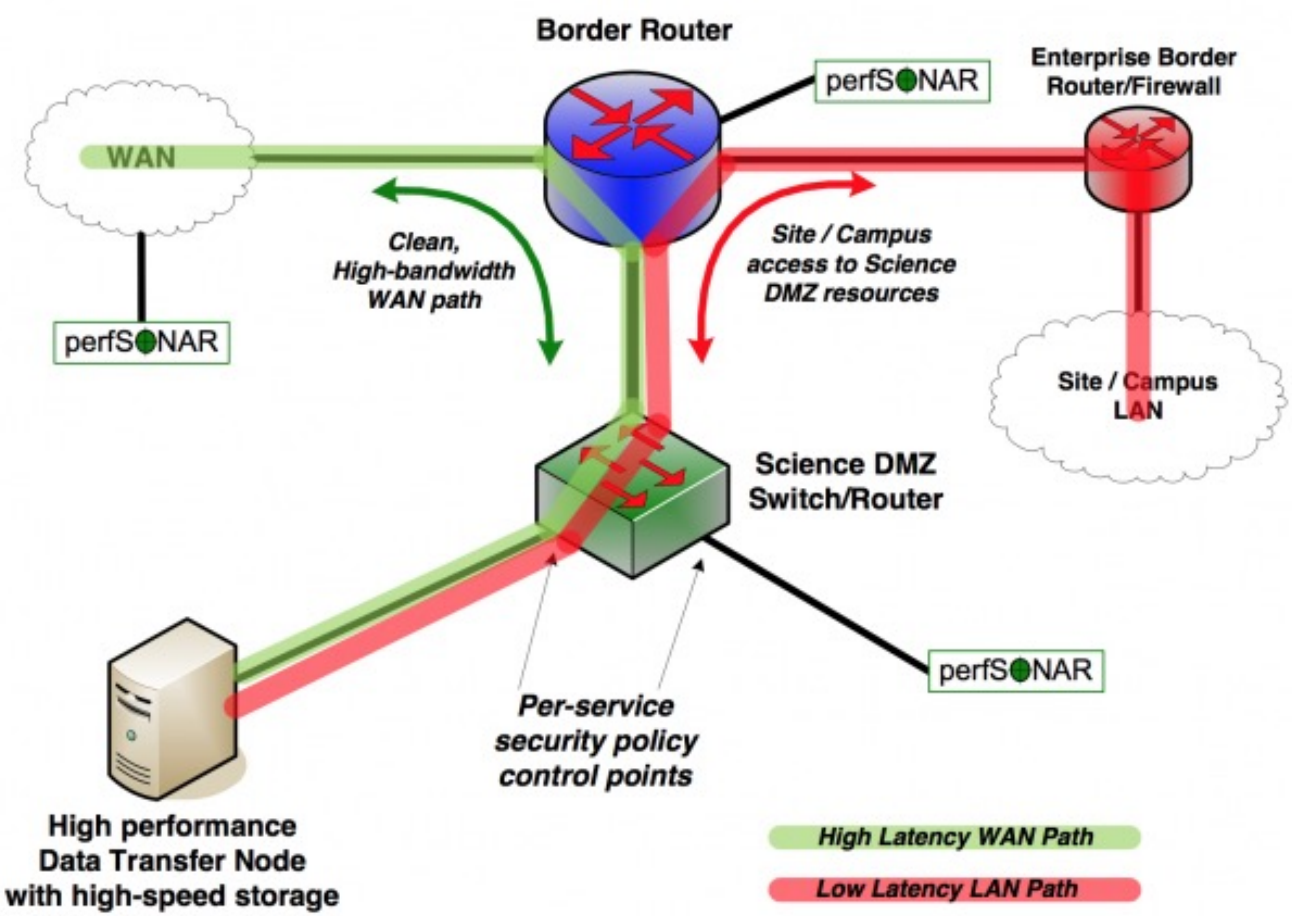
- An area of interest for the experiments is traffic pacing/shaping.
  - Without traffic pacing, network packets are emitted by the network interface in bursts, corresponding to the wire speed of the interface.
    - Problem: microbursts of packets can cause buffer overflows
    - The impact on TCP throughput, especially for high-bandwidth transfers on long network paths can be significant.

Emerging Networking R&D – Network Monitoring & AI/ML for Networking

- Network Monitoring:
  - To design and develop a comprehensive network monitoring platform for HEP community using various monitoring systems (e.g. perfsonar)
    - Provides data and feedback to LHCOPN/LHCONE, HEPiX, WLCG and OSG communities
- AI/ML for Networking:
  - The scientific network traffic patterns are not well understood, or at best, understood only at a very high level
  - Network traffic patterns need to be investigated and characterized at different granularity levels
  - To exploit the various AI/ML technique to train and evaluate the traffic classification models by using Machine Learning/Deep Learning algorithms.
  - One example of traffic patterns: file transfer vs streaming traffic

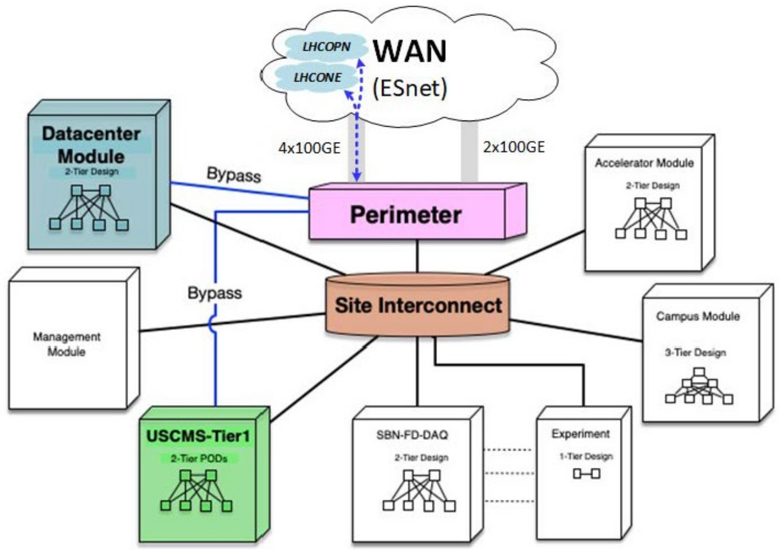
Additional Slides

Science DMZ Architecture – Conceptual Diagram



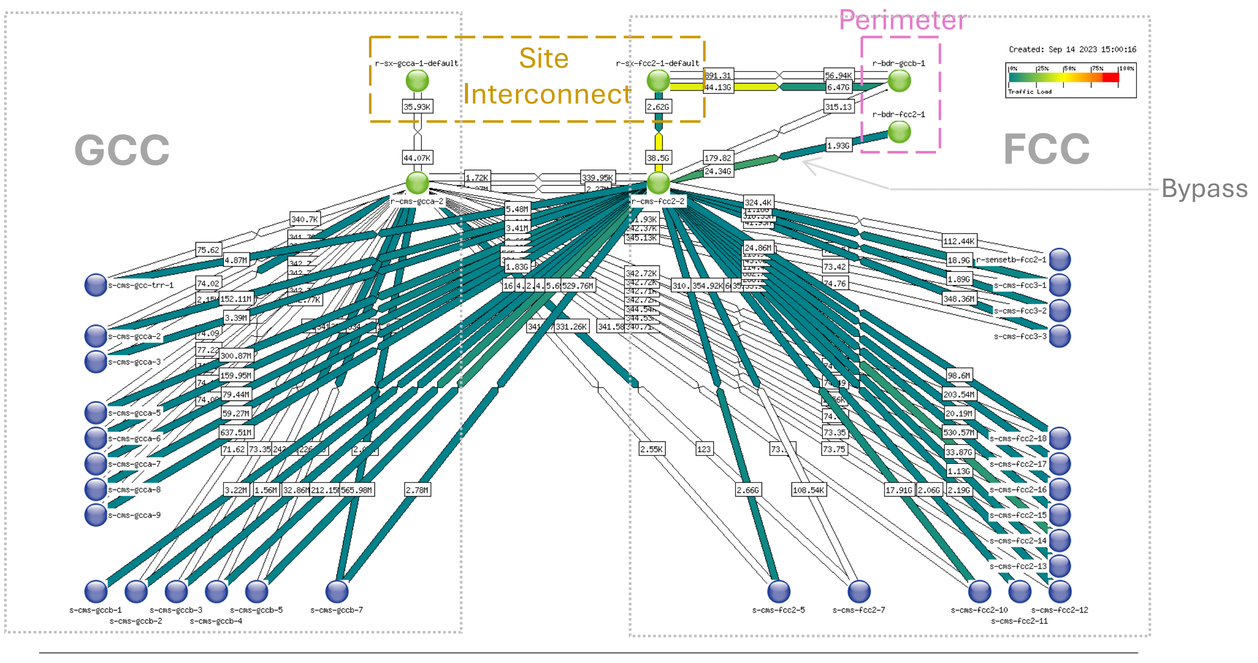
FNAL Network Architecture – High-Level

- Modular design:
  - DataCenter for general data center computing resources
  - USCMS-Tier1 has its own module
  - Site Interconnect module for intra-site & off-site connectivity
  - Perimeter module delineates & controls off-site connectivity
  - USCMS-Tier1 & Datacenter modules have direct connection to perimeter for science network traffic



U.S. CMS Tier1 Module at FNAL

- Distribution switches in FCC2 & GCC-A
  - Also large, modular devices in dedicated racks
- ToR access switches dual-homed to distribution switches:
  - Large-scale deployment of 100GE host connections



Thank you ☺

Any question?