Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

ePIC Network Architecture and Connectivity Options

Mark Lukasczyk
Network Engineering

@BrookhavenLab

10/8/2025

BNL and JLAB Connectivity via ESnet

Path latency between BNL and JLAB:

NYC to Washington DC

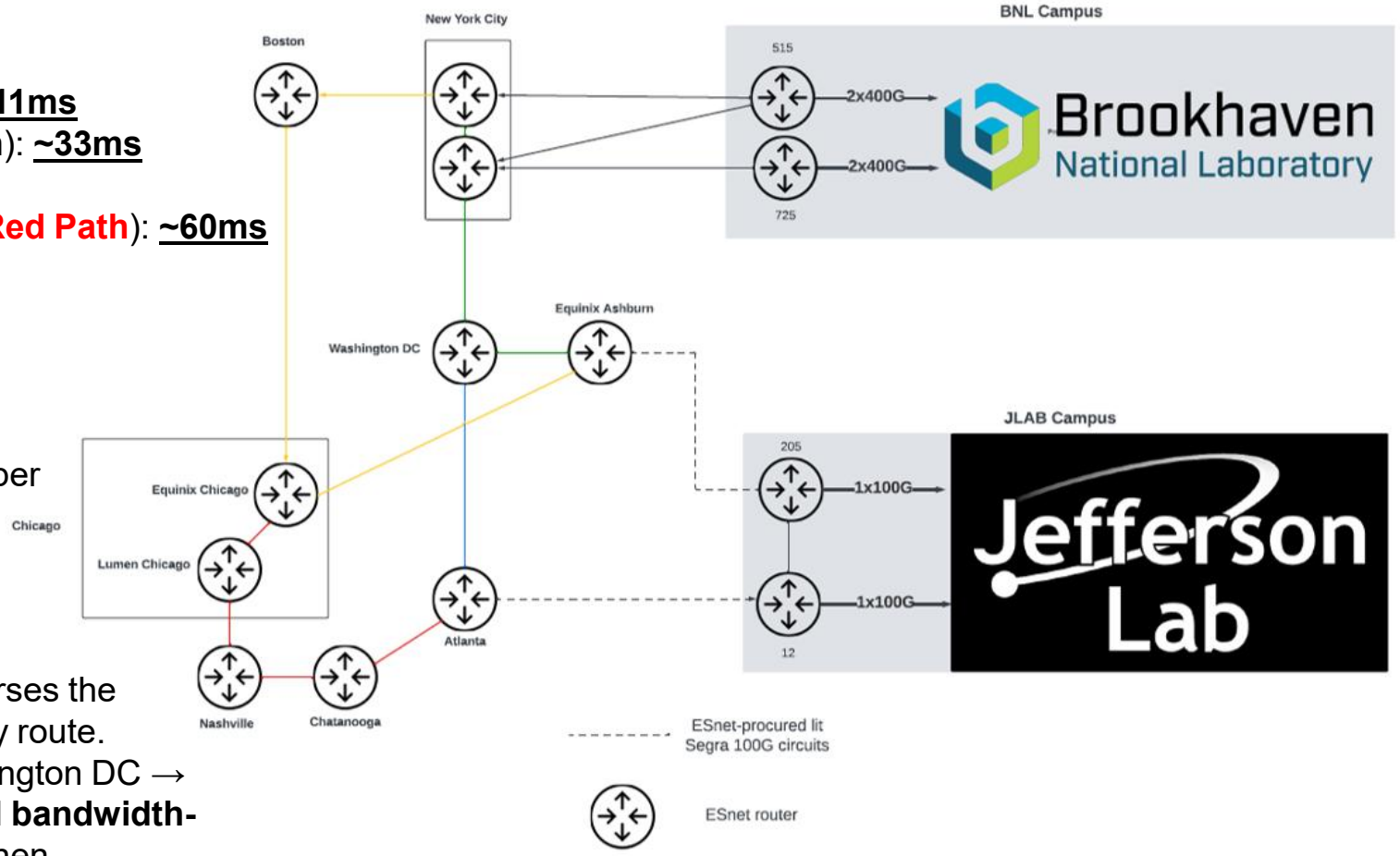
- NYC → Washington DC → Ashburn (**Green Path**): **~11ms**
- NYC → Washington DC → Atlanta (**Green/Blue Path**): **~33ms**

NYC to Boston (Diverse Path):

- Boston → Chicago → Nashville → Atlanta (**Orange/Red Path**): **~60ms**

Network Overview:

- ESnet backbone is built on Nx400G optical links.
- The current constraint is the last mile to JLAB, which is limited to two 100G Segra-procured circuits.
- ESnet is evaluating the feasibility of deploying dark fiber to JLAB, though this remains under investigation.

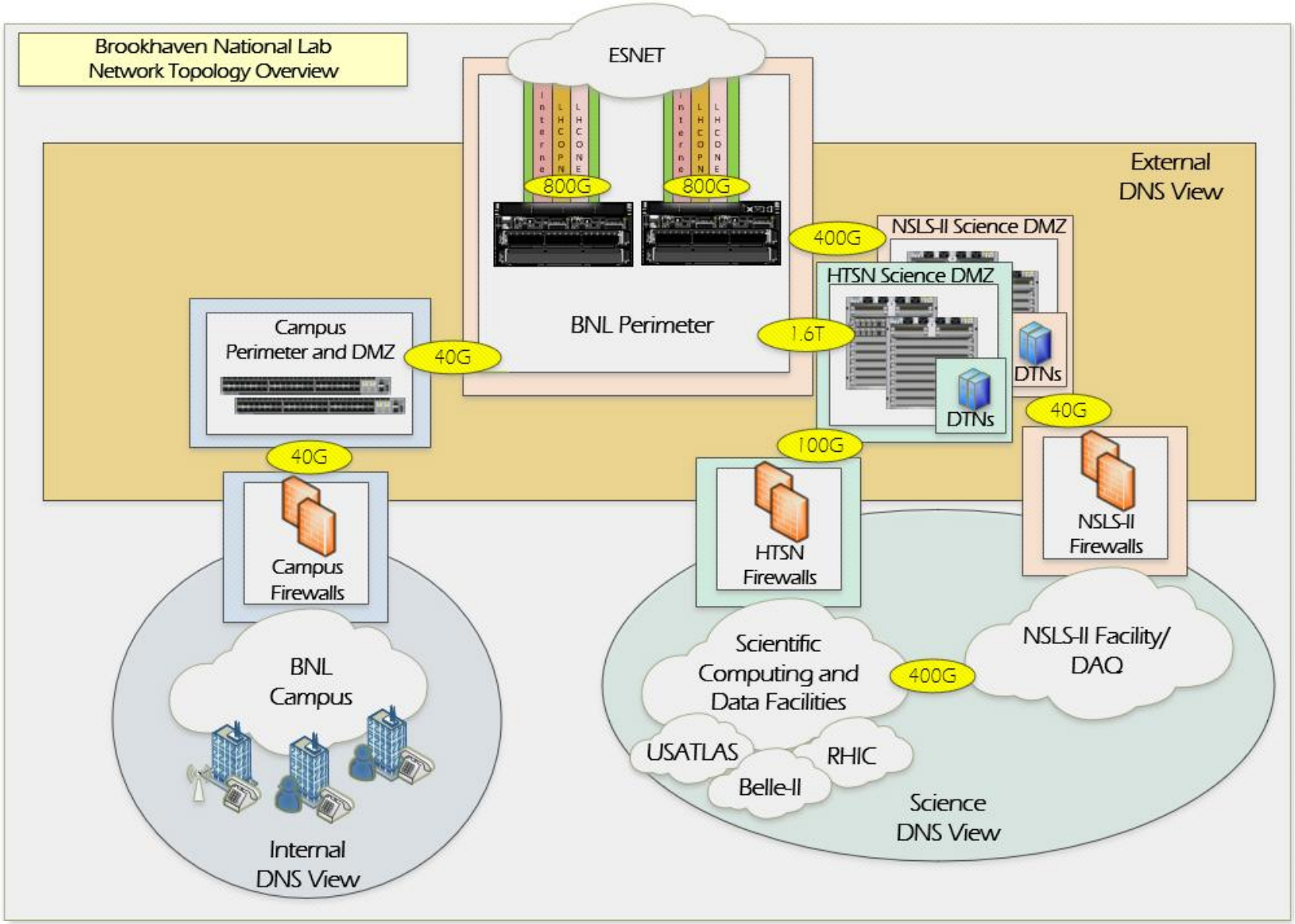


Current Operations and Considerations:

- Today, traffic between BNL and JLAB primarily traverses the **green path**, selected as the best path/lowest-latency route.
- Since the primary path to JLAB is via NYC → Washington DC → Ashburn, any virtual circuits must be **prioritized and bandwidth-guaranteed** to ensure reliable performance, even when competing with other collaborations (e.g., LHCONE/LHCOPN).

Brookhaven National Laboratory

BNL High Throughput Science Network (HTSN)



Brookhaven National Laboratory

ePIC Internet/WAN Network Topology Option 1 - Most Preferred

- The proposed topology closely mirrors the LHC model, where VPLS/OSCARS circuits terminate on BNL's network perimeter and extend point-to-point connectivity with JLAB via EBGP.

- **Primary connectivity:** Dedicated VPLS/OSCARS circuits between BNL and JLAB (guaranteed bandwidth, predictable).
- **Backup connectivity:** ESnet backbone provides resilient failover to JLAB or other destinations if VPLS/OSCARS circuits are unavailable.

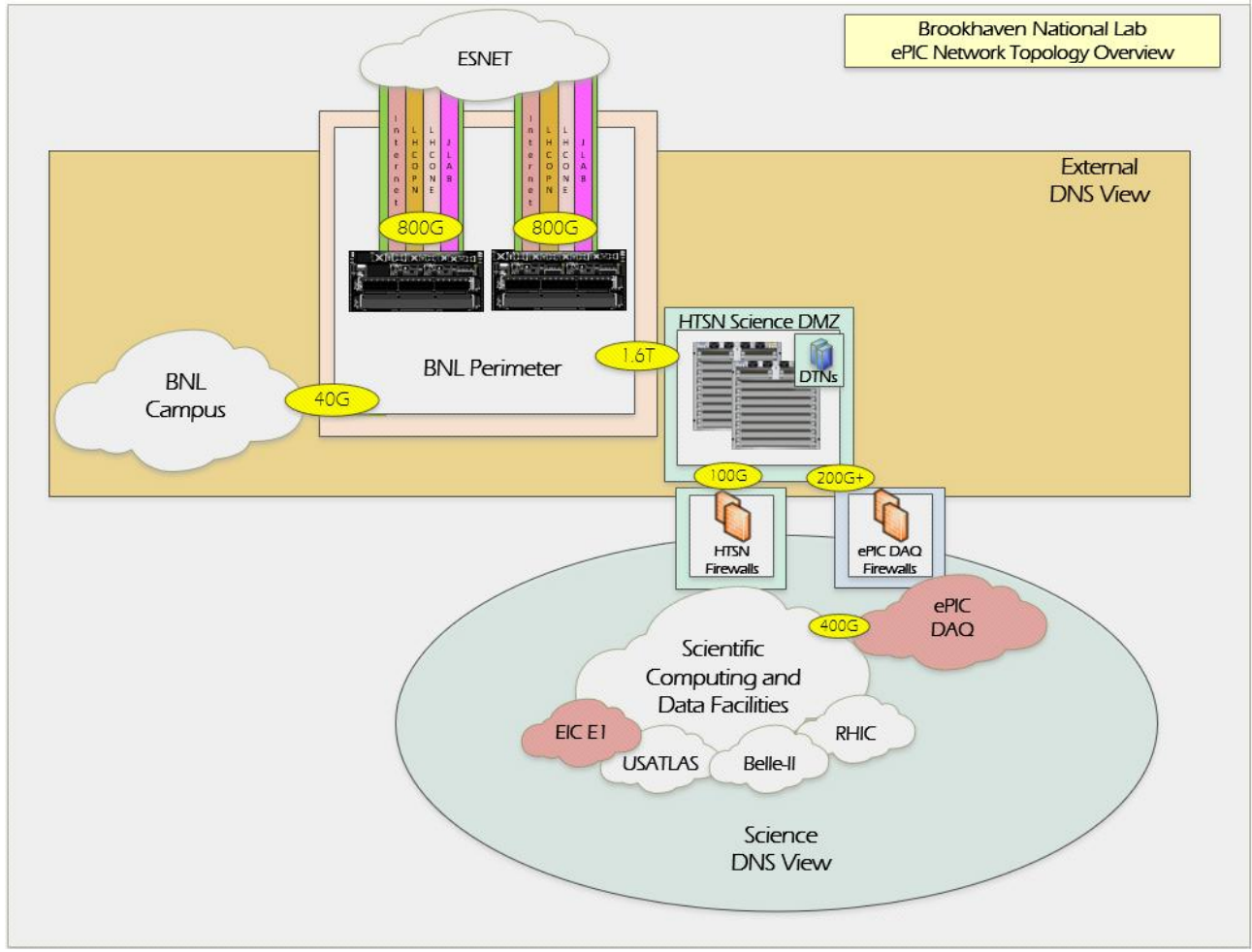
- Provides connectivity for both the **ePIC DAQ (E0)** and **BNL E1** to transfer data to JLAB.

- No new physical circuits are required from ESnet.

- Maintains a single point of entry into BNL from ESnet, enabling Cyber Security and Network Engineering to leverage existing monitoring and mitigation infrastructure.

- Introduces a high-throughput firewall between the ePIC DAQ, the Internet, and JLAB, ensuring controlled access and failover routing.

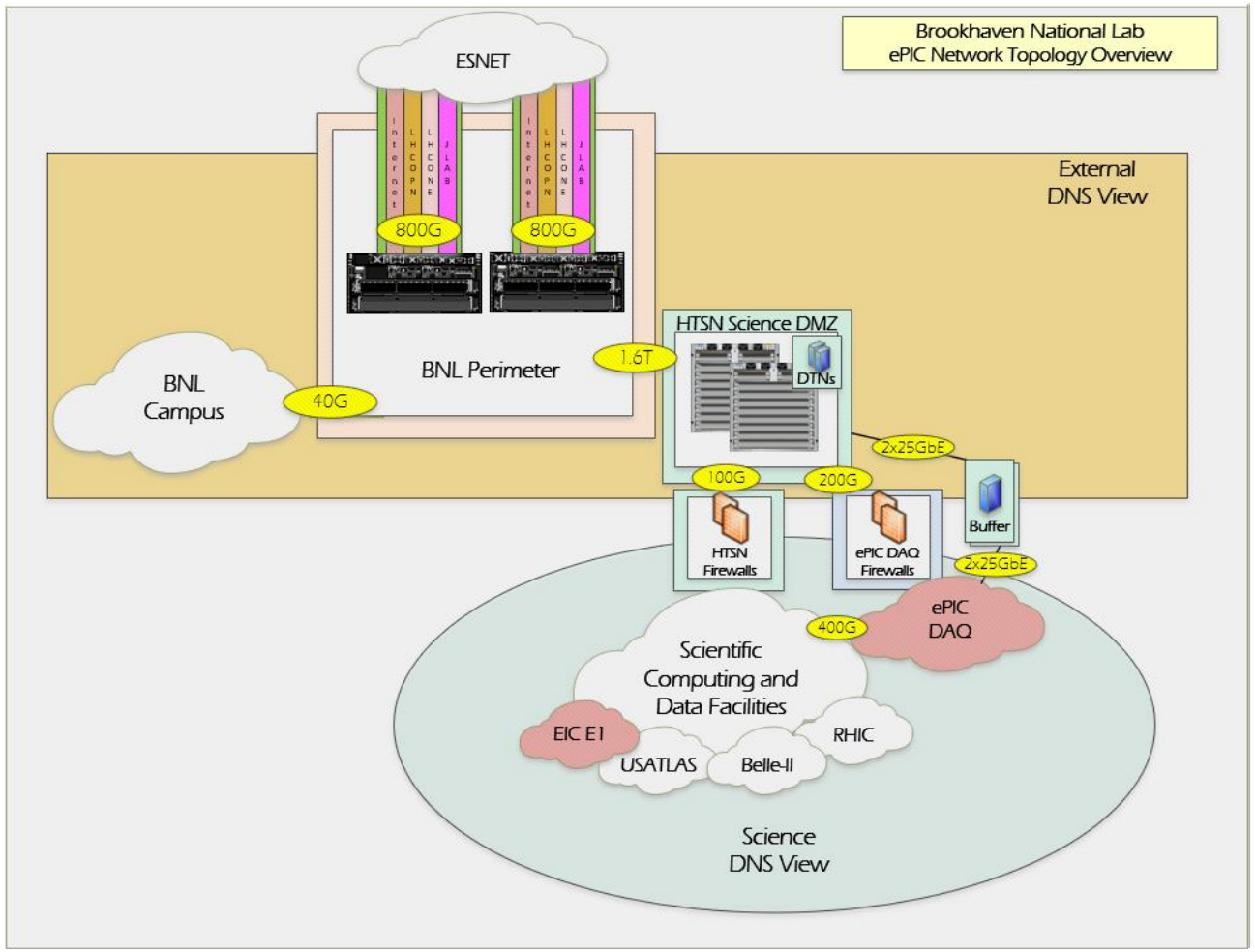
- Connects the ePIC DAQ to the SDCC via the HTSN Science Core, using route filtering and ACLs.



Brookhaven National Laboratory

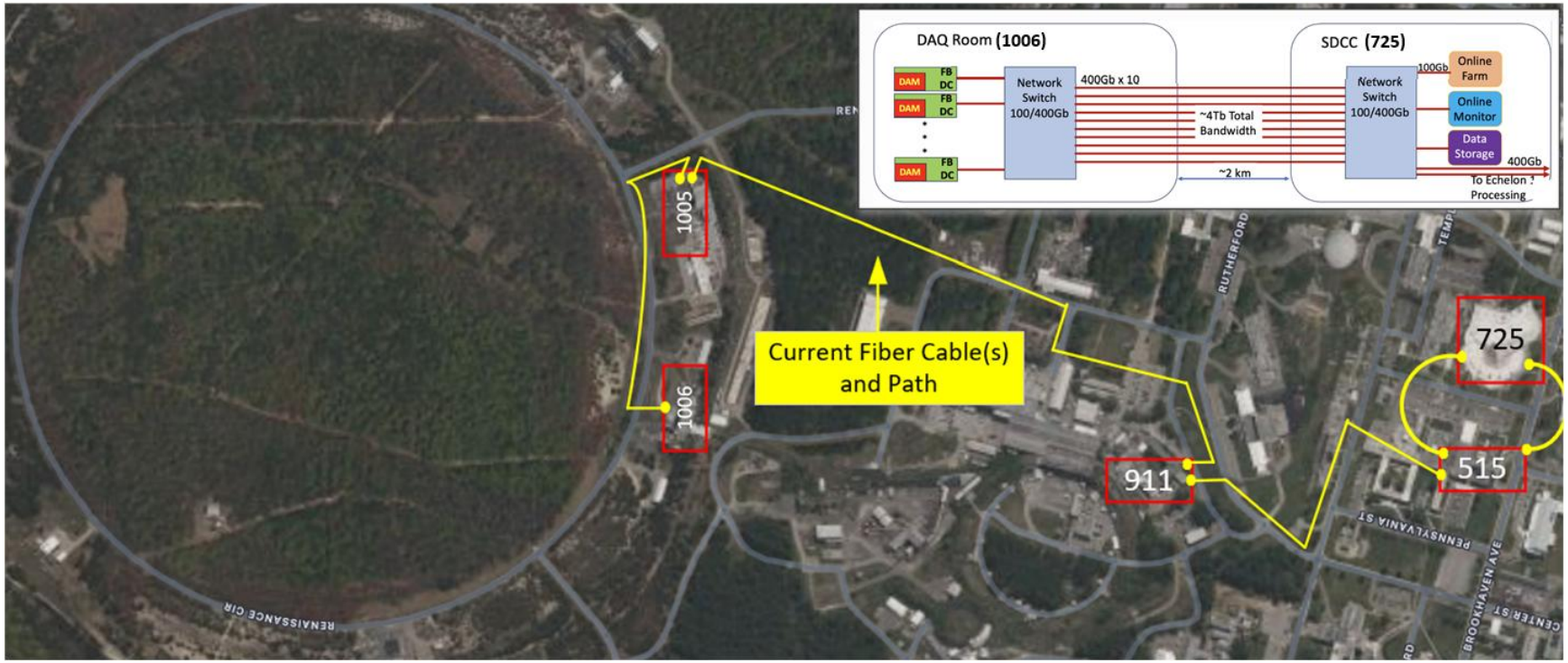
ePIC Internet/WAN Network Topology Option 2

- This option is **similar to Option 1**, except that if placing all E0 resources behind a firewall is not **technically feasible**, the buffer can instead be multi-homed to the BNL Science DMZ.
- Buffer boxes can be placed on their own subnet, with local iptables rules providing basic access controls.
- Because the primary data path between E0 and JLAB bypasses the firewall, a lower-performance firewall could be used for supporting functions.
- Tradeoffs:
  - The buffer is **not protected by a firewall**, requiring system administrators to harden the servers directly.
  - Granular connectivity controls and centralized logging available with a firewall would be lost.
  - Routing complexity increases, as administrators must carefully define and maintain routes on the multi-homed buffer systems.



Brookhaven National Laboratory

Fiber Path Between IP6 (1006) and SCDF (725)

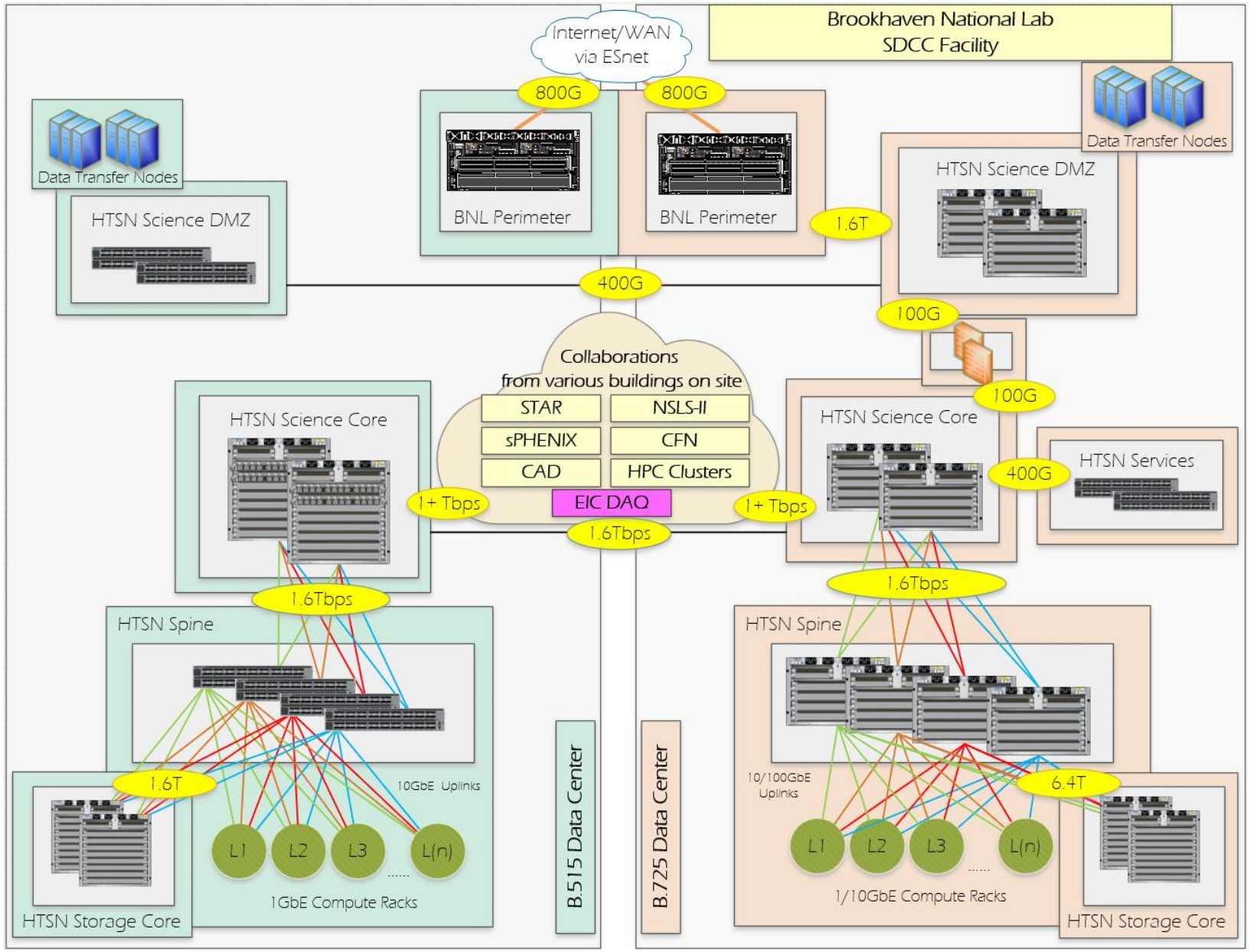


- Consists of a series of single multi-fiber cable runs, shown in yellow, connecting buildings 515, 911, 1005, 1006.
- The cables are terminated in patch panels in each of the buildings, and patch cables are used to connect the fibers between the cables.
- A break in any one of these cables would result in loss of connectivity from ePIC to the data center.
- Currently, approximately two-thirds of the fiber capacity between buildings 515 and 1005 has been utilized in support of RHIC.

Questions

Backup Diagrams/Slides

HTSN/SDCC Network Infrastructure

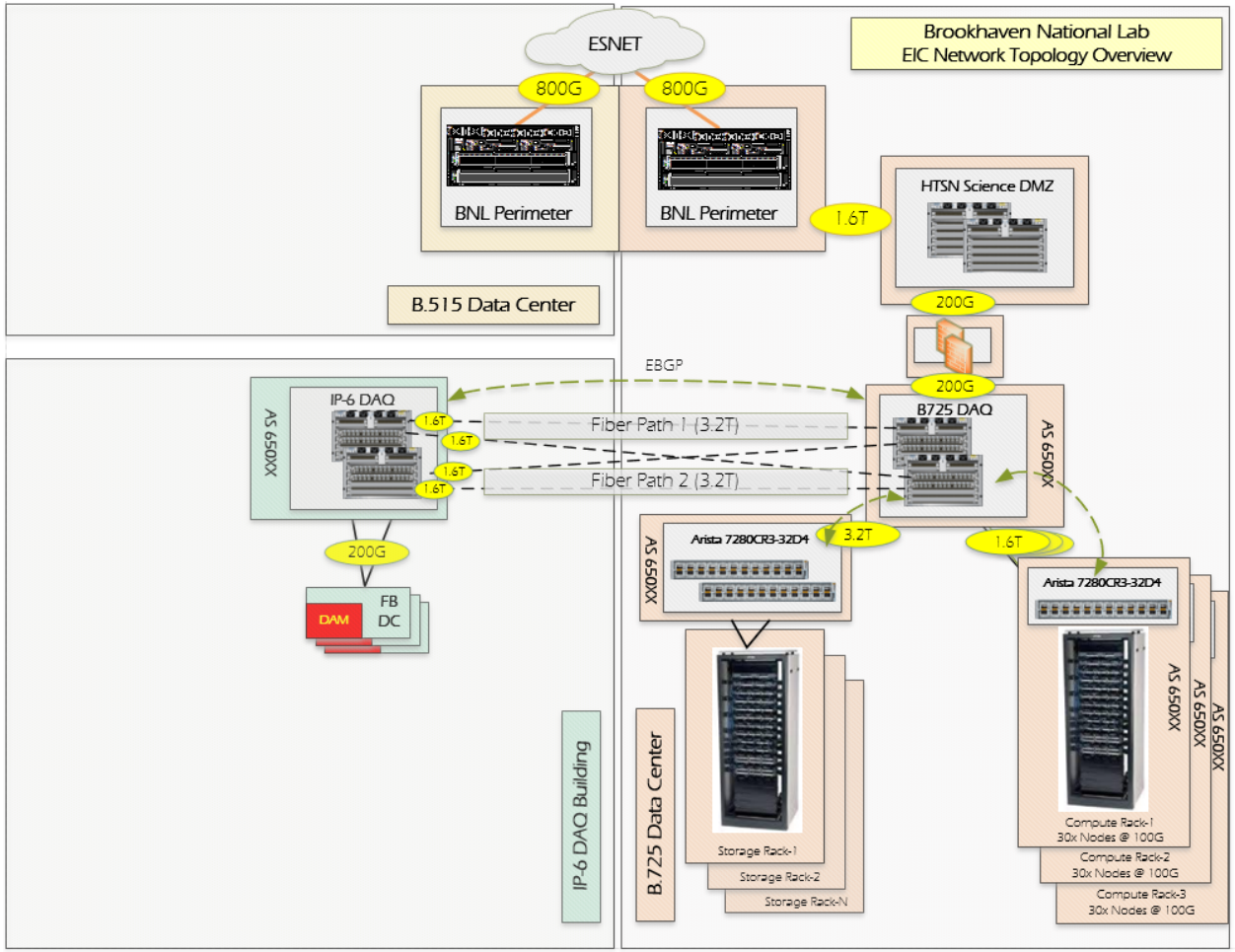


Brookhaven National Laboratory

Internal ePIC DAQ Network Infrastructure – Option 1

- **IP-6 DAQ Network Infrastructure**
  - Could consist of two 4 slot Arista chassis' with:
    - 2x 24 port 400GbE line cards.
    - 2-4x 36 port 100GbE line cards.
    - ~3km from building 1006 to 515 on current fiber path. Will require the use of LR transceivers.

- **Building 725 DAQ Network Infrastructure**
  - Spine could consist of two 4 slot Arista 7804 chassis' with:
    - 2x 36 port 400GbE line cards.
  - Compute racks should utilize Top of Rack (ToR) Arista 7280CR3-32D4 or similar switches:
    - Switches provide 32x100GbE and 4x400GbE
    - This would allow us to connect the compute ToRs in any location within the 725 MDH or even external locations.
  - Storage head nodes will be redundantly connected to a centralized MLAG pair of Arista 7280CR3-32D4 or similar switches.
    - If the storage head nodes are 25GbE we may want to use a different type of fixed chassis switch.
    - Head nodes will be connected via fiber cabling back to the centralized Arista switches. This will allow the storage racks to be located anywhere within the 725 MDH.

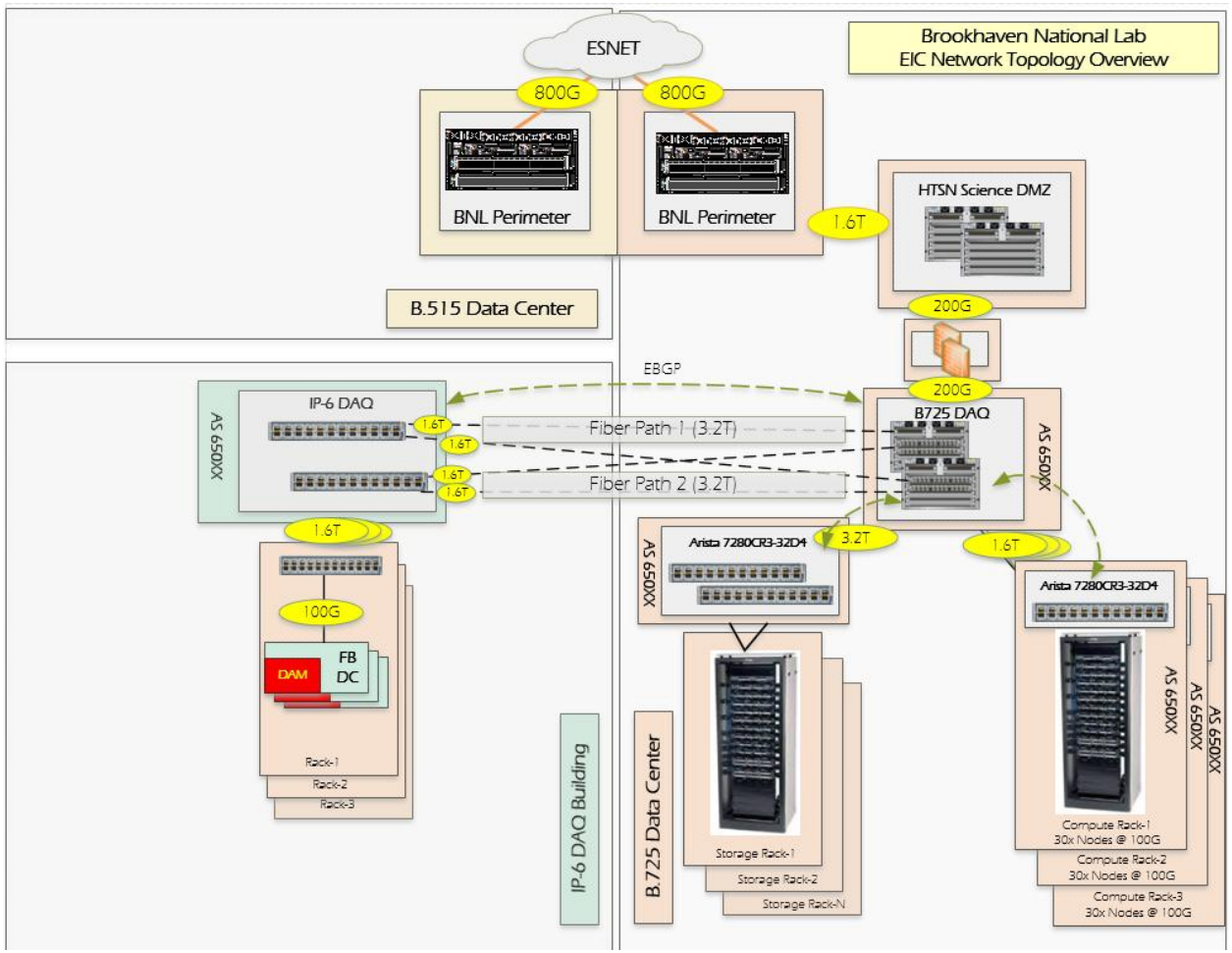


Brookhaven National Laboratory

Internal ePIC DAQ Network Infrastructure – Option 2

- **IP-6 DAQ Network Infrastructure**
  - Two Arista fixed chassis switches each supporting 24 port 400GbE.
  - Each rack containing DAM servers will receive a ToR Arista 7280CR3-32D4 (32x100GbE and 4x400GbE)

- **Building 725 DAQ Network Infrastructure (same as option 1)**
  - Spine could consist of two 4 slot Arista 7804 chassis' with:
    - 2x 36 port 400GbE line cards.
  - Compute racks should utilize Top of Rack (ToR) Arista 7280CR3-32D4 or similar switches:
    - Switches provide 32x100GbE and 4x400GbE
    - This would allow us to connect the compute ToRs in any location within the 725 MDH or even external locations.
  - Storage head nodes will be redundantly connected to a centralized MLAG pair of Arista 7280CR3-32D4 or similar switches.
    - If the storage head nodes are 25GbE we may want to use a different type of fixed chassis switch.
    - Head nodes will be connected via fiber cabling back to the centralized Arista switches. This will allow the storage racks to be located anywhere within the 725 MDH.



Brookhaven National Laboratory