

The Data-Intensive Network

Guy Almes and Willis Marti
Texas A&M University

Cyberinfrastructure

- Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable breakthroughs not otherwise possible.



- Typical Elements
 - TeraGrid HPC clusters
 - Campus 200-TByte Storage System
 - Local Visualization Studio
 - Remote Doppler Weather Radar
- Can these play together?
- Can we craft workflows using these resources in flexible ways?

Networking Requirements

- Very high performance flows among serious cyberinfrastructure resources
- Relatively small number of relatively heavy flows
- Predictable availability
- Predictable latency
- (non-requirement: firewall)

Where does Internet2 fit in?

- Take a “retro” definition of Internet2
 - combined backbone, RON, campus net
 - this end-to-end network is the only network that matters for science users
- Where is the bottleneck between cyberinfrastructure resources on different campuses?
 - 1988 answer
 - 2009 answer

Current Network Environment

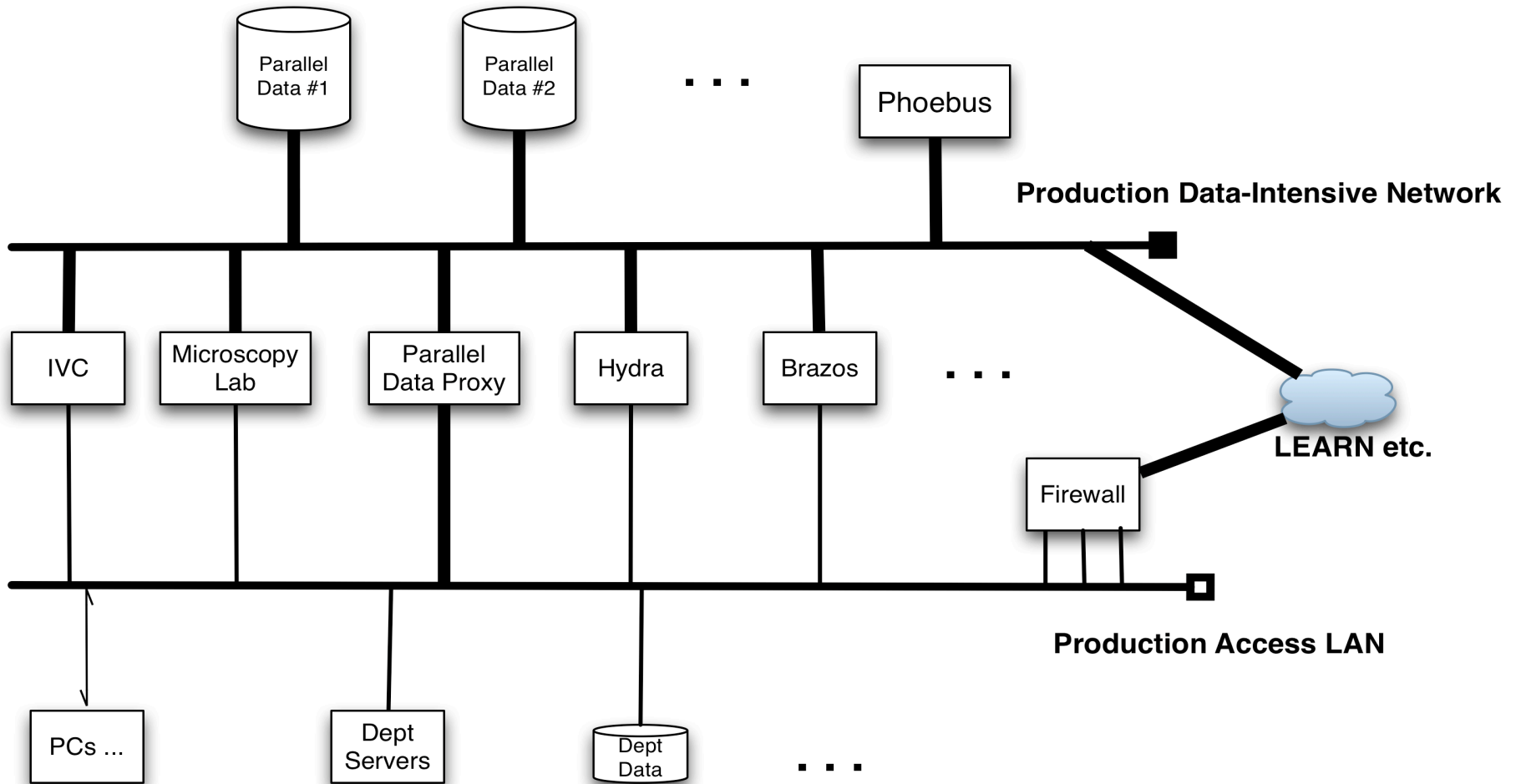
- Backbone(s) and RONS
 - 10 Gb/s IP
 - 1-10 Gb/s (dynamic) circuits
 - DWDM fiber plants
- Campus Access LANs
 - 100-Mb/s to 10-Gb/s IP
 - “traffic disruption devices”
 - huge number of devices to manage




Proposed solution

Data Intensive Network

- Separate this DIN from the Campus Access Network
- Resources needing access to both networks can be dual-homed
- Provide security by monitoring and cooperation

Texas A&M University Data-Intensive Network



-  10/100 Mb/s
-  10 gigE
-  gigE

Site Influences

- Distributed solution
 - 5,200 acres
 - 340+ buildings
 - Not fiber rich
- Use several small-ish high performance switches
 - Testing Cisco Nexus 5000
 - Looked at Extreme, Foundry, Force10

Site Selection

- Two primary data centers
 - Includes connection to LEARN
- Key cluster users
 - Chemistry
 - Geosciences
 - Physics (new building)
 - Life Sciences (new building)
- Immersive Visualization Center
- Microscopy Lab (shared instruments)

Implementation Notes

- Separating campus access and data-intensive networks allows optimization of each
- Separate, better, fewer, but more costly switches
- Security through host and traffic monitoring, not through packet-forwarding devices

More Implementation Notes

- DIN allows shared special services:
 - Phoebus gateway
 - Dynamic Circuit Network interface
 - Performance measurement infrastructure