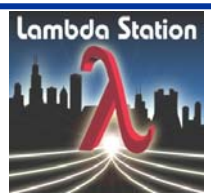


TeraPaths

Phoebus

Network Middleware: Lambda Station, TeraPaths, Phoebus

Matt Crawford

GLIF Meeting; Seattle, Washington

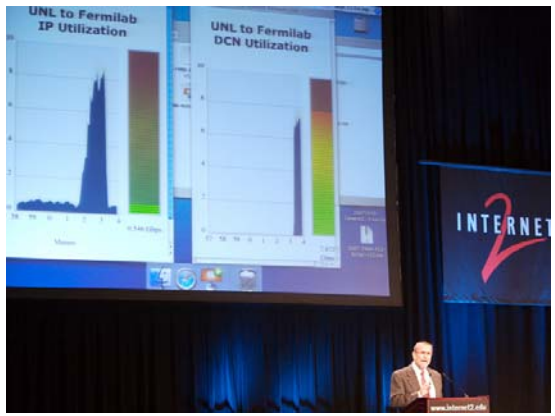
October 1-2, 2008

Office of Science

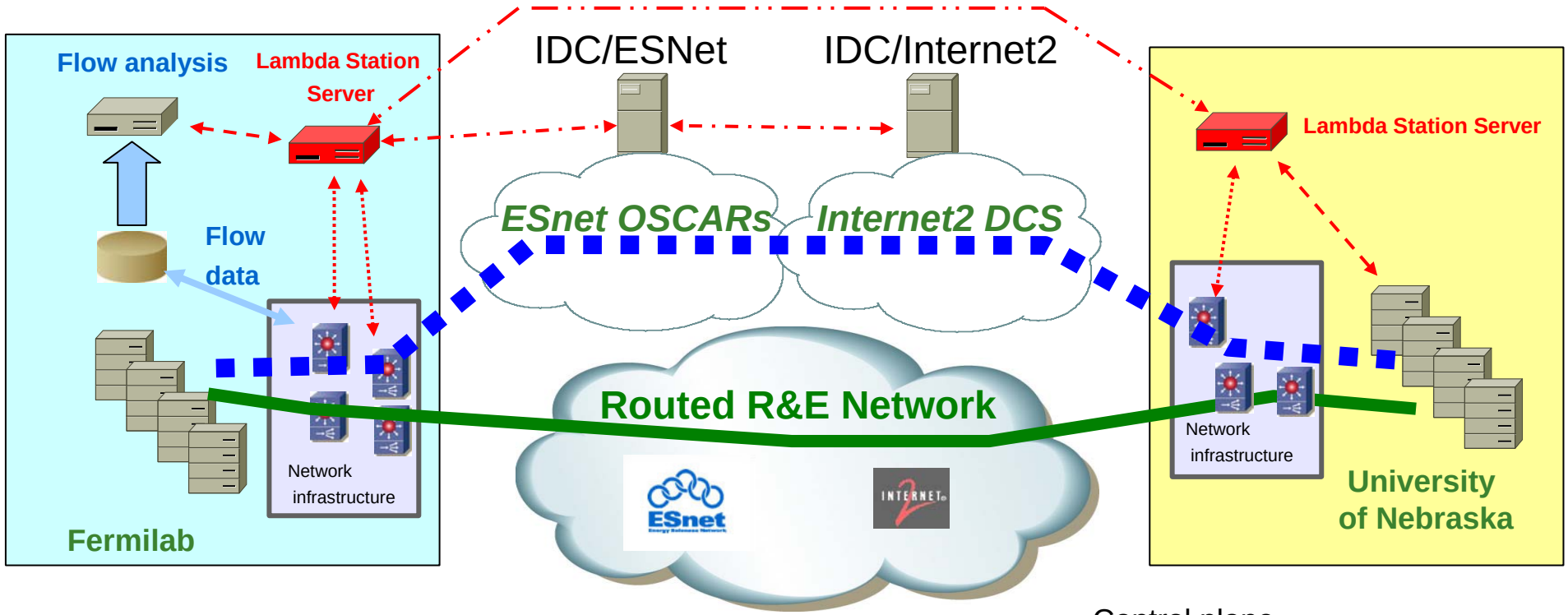
U.S. DEPARTMENT OF ENERGY

Lambda Station (I)

- Target: last-mile problem between local computing resources and alternate WAN paths
- Strategy - develop a network path selection service
  - Setup & teardown alternate (dynamic) WAN paths
  - Reroute local traffic over alternate network paths
    - Selective forwarding on source/destination netblocks
  - Graceful cutover & fallback
  - On-demand from applications (SRM)
    - User-initiated as well
    - Flow data trigger, too
- Practical deployment → LHC data movement



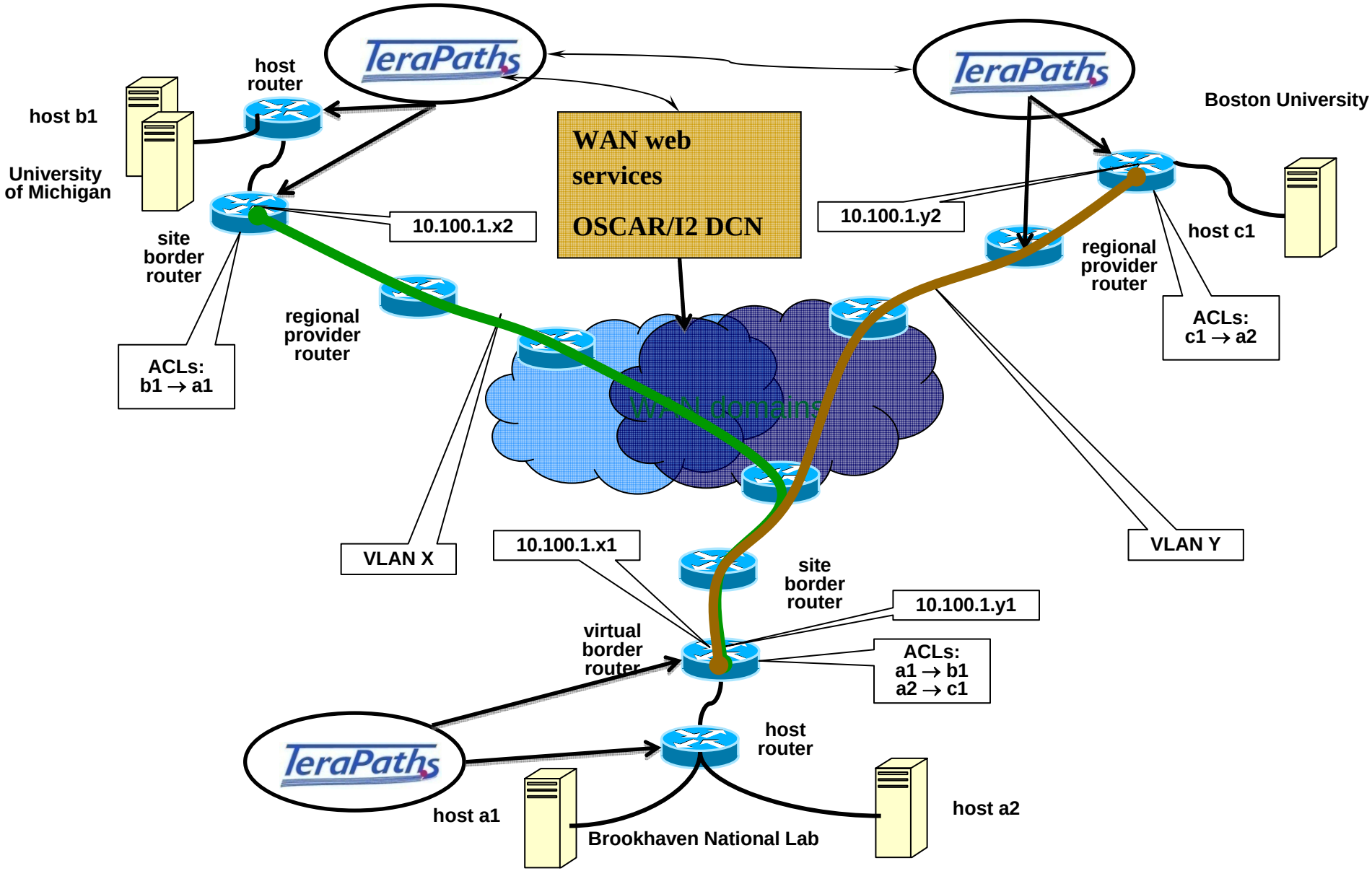
Lambda Station (II)



TeraPaths Introduction

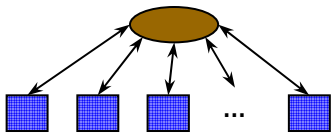
- Goal: Support efficient, predicable, and prioritized peta-scale data replication in modern high-speed networks
- Movitation:
  - Data flows have varying priority/importance
    - Video streams
    - Critical data
    - Long duration transfers
  - Default “best effort” network behavior treats all data flows as equal
  - Capacity is not unlimited
    - Congestion causes bandwidth and latency variations
    - Performance and service disruption problems, unpredictability
- Solutions:
  - Establishes on-demand and manages true end-to-end, QoS-aware, virtual network paths across multiple administrative network domains
  - Dedicates network resources to data flows specifically authorized to use such network paths, in a transparent and scalable manner. This ensures that only selected flows receive a pre-determined, guaranteed level of QoS in terms of bandwidth, jitter, delay, etc.
  - Integrates into Data Transfer Tools (SRM)
    - Flexible Plug-in to allow dynamic reservation to TeraPaths

TeraPaths End-to-End Setup

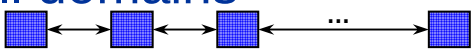


Establishing End-to-End QoS Paths

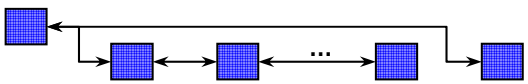
- Multiple administrative domains
  - Cooperation, trust, but each maintains full control
  - Heterogeneous environment
  - Domain controller coordination through web services
- Coordination models
  - Star
    - Requires extensive information for all domains



  - Daisy chain
    - Requires common flexible protocol across all domains



  - Hybrid (end-sites first)
    - Independent protocols
    - Direct end site negotiation

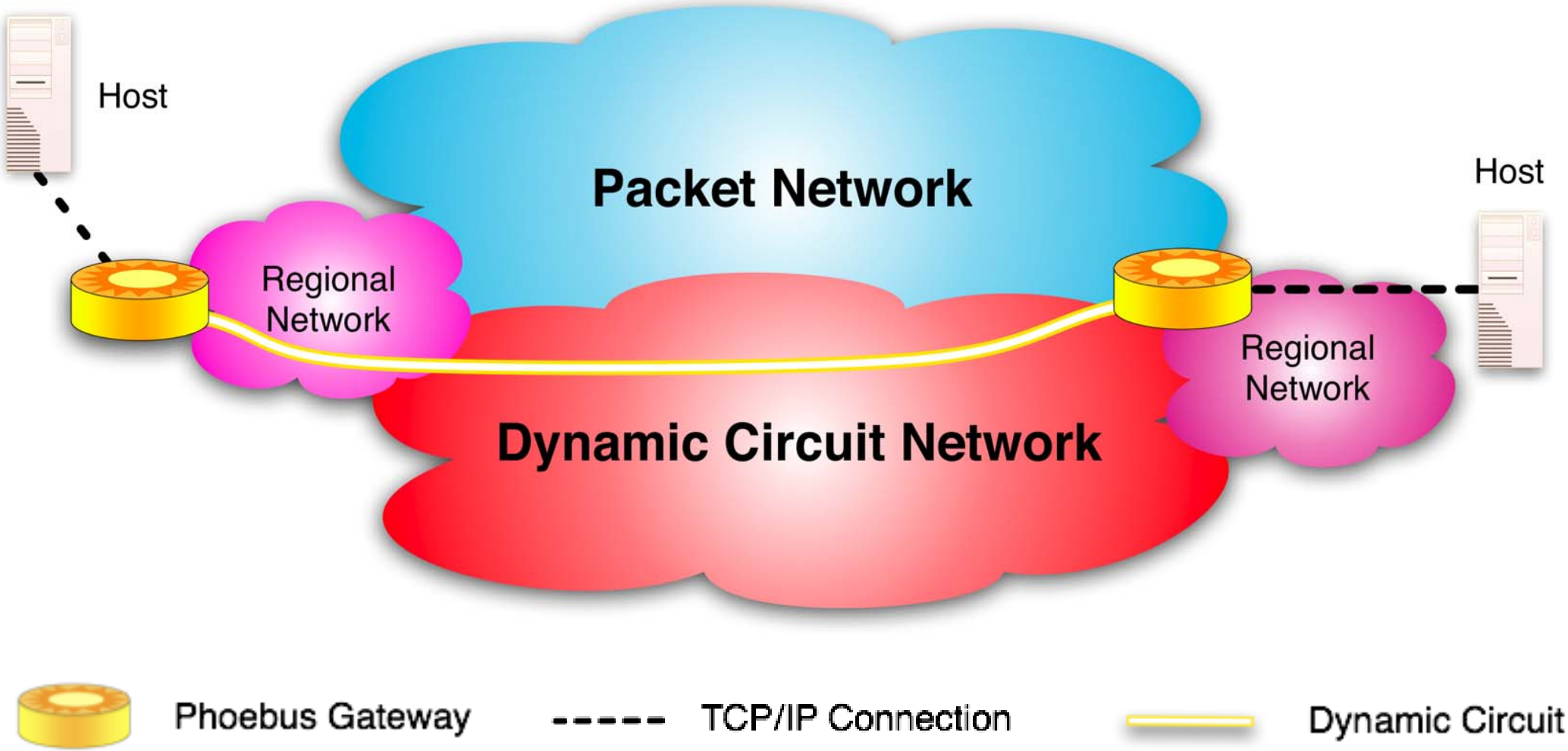


Phoebus

- Goal: High performance interface to hybrid networks with low barrier for entry for users

- Solution:
  - Phoebus Gateways installed at the interface between circuit and packet networks as a transparent “on ramp”
  - Session layer interface allows the path to be segmented into various transport-layer hops
  - Manages circuit allocation and aids Transport-layer performance

Phoebus

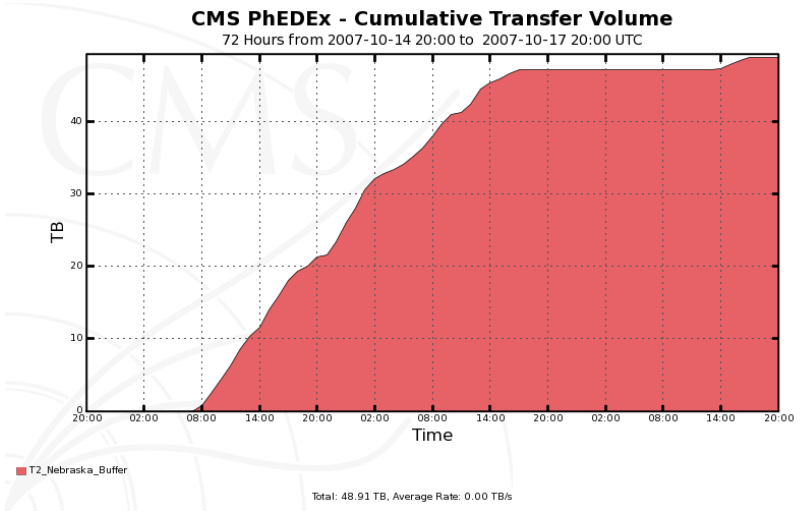

Host
Packet Network
Host
Regional Network
Regional Network
Dynamic Circuit Network
Phoebus Gateway
TCP/IP Connection
Dynamic Circuit

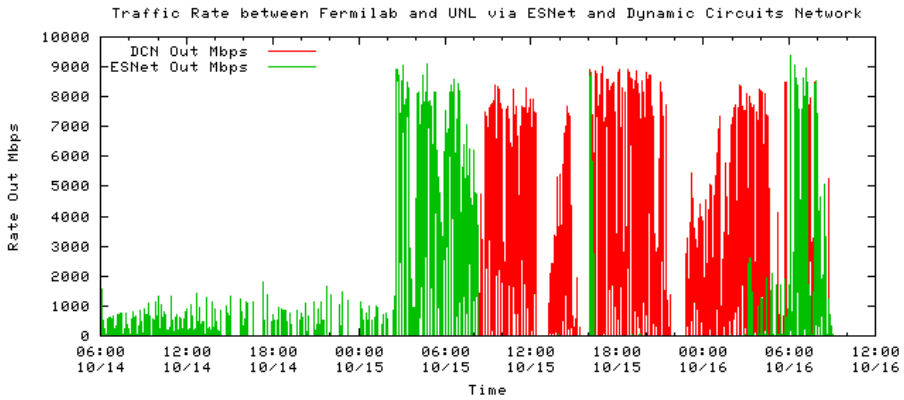
Network middleware

- All three products function as “network middleware”
  - Interface to applications via plug-in: I.e. dCache SRM Plugin for TeraPaths and LambdaStation is committed into dCache release
  - Interface to network infrastructure: Map network requests into the network devise configuration.
  - Interface to other Network middlewares: coordinates with each other to set up end-to-end network paths.
- Services that improves quality of network performance for applications
- Important element(s) in use of emerging dynamic circuit services

Current Deployment – Lambda Station

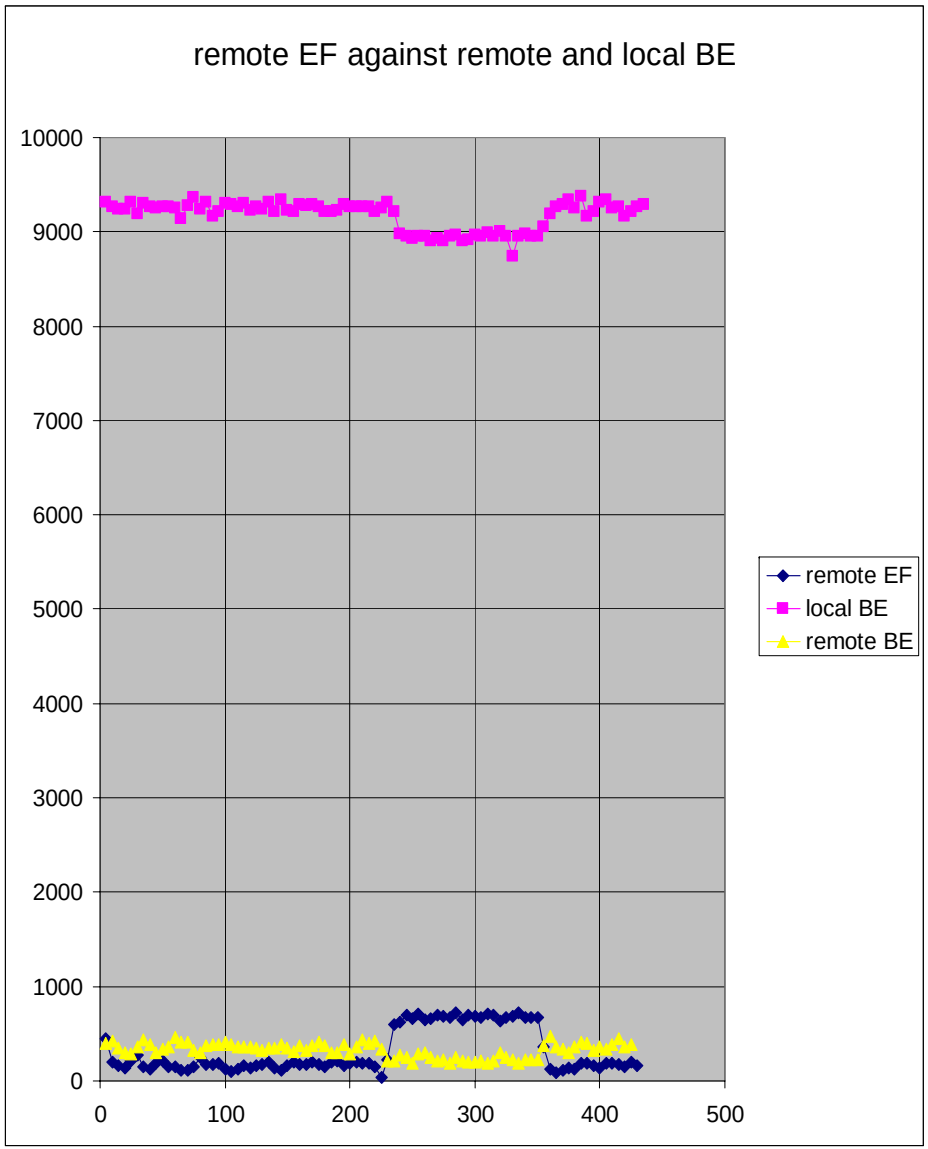
- **Deployed at FNAL, Caltech, Nebraska (UNL)**
  - In ***production*** use for CMS Tier1/Tier2 data movement
- **Large-scale data recovery via DCN in October, 2007**
  - UNL loses their Tier-2 data cache
  - 50TB cache recovered by transfer from FNAL Tier-1:
    - ⇨ Largely via Internet2/ESNet Dynamic Circuit
    - ⇨ Completed in 32 hours





Current Deployment - TeraPaths

- **Deployed at BNL, University of Michigan, Boston Univ.**
  - In use for USATLAS Tier2/Tier2 data movement
- **Effectiveness of TeraPaths was demonstrated at SC2007 for mutual protection among competing flows**
- **Large-scale data transfer via I2 DCN in Sep, 2008**
  - BU and Univ. of Mich. share their Tier-2 data caches
  - A Tier 2 can pull data from other Tier 2 when BNL is not accessible.
  - Largely via Internet2 Dynamic Circuits



Current Deployment - Phoebus

- Internet2 has installed a prototype Phoebus infrastructure on its backbone network and is investigating offering a production service to its members
- NYSERNet, New York's Regional Optical Network, has deployed and is testing for use with LIGO data transfers
- ESnet has been experimenting with Phoebus
- GEANT2 has also utilized Phoebus to enable dynamic bandwidth allocation (with the DCN-like AutoBAHN) and in support of the Electronic Very Long Baseline Interferometry (eVLBI) effort
- RNP in Brazil and KiSTi in Korea are also experimenting with Phoebus

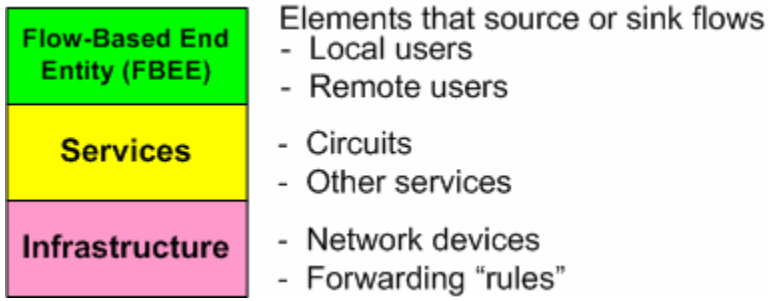
Convergence

- Emerging dynamic circuit services are:
  - Lambda Station's alternate WAN paths
  - TeraPath's end-to-end QoS WAN paths. (Multi-Layer: layer 2 VLAN and layer 3 MPLS)
  - Phoebus Gateway-to-Gateway paths
- Products end up serving the same general functions:
  - Coordinate with WAN inter-domain controllers
  - Configure site network for alternate ingress/egress points
  - Provision network bandwidth to individual flows
  - End to end network path creation and management
  - Sharing common user community (LHC: ATLAS and CMS)
  - Coordinated projects will accelerate technology evolution
- New short-term (3-6 month) goal – interoperate!
  - We are investigating opportunities & obstacles

Coordinating with related efforts

- Conforming to circuit “standards” is a high priority
- But who & what are they?
  - Network interface standards (DCN/OSCARs, Autobahn, etc)
  - OGF Network Services Interface (NSI) & Network Markup Language (NML) working groups
  - User/application interface standards
  - What else?
- We see a need for abstract functional representations
  - Topology representation model?
  - Common reference model?

Lambda Station/TeraPaths Reference Model →



Future Directions

- Early design study on End Site Domain Control Service
- Envisioned as end component of end-to-end cross-domain control plane infrastructure
  - True end-to-end, not just to site perimeter
- Would be built upon Lambda Station, TeraPaths, Phoebus development & experiences
- Additional capabilities might include:
  - E2E path monitoring status
  - Performance analysis of achieved application performance
- Practical application (vetting) on LHC data movement
  - Integration with LHC Data Management Stacks:
    - PHEDEX, FTS, dCache/SRM: CMS
    - ATLAS DDM (Distributed Data Management), FTS, and SRM.