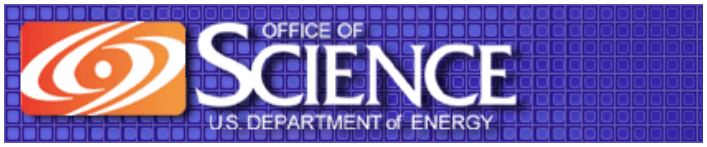


High Performance Networks for High Impact Science

Report on the
High Performance Network Planning Workshop

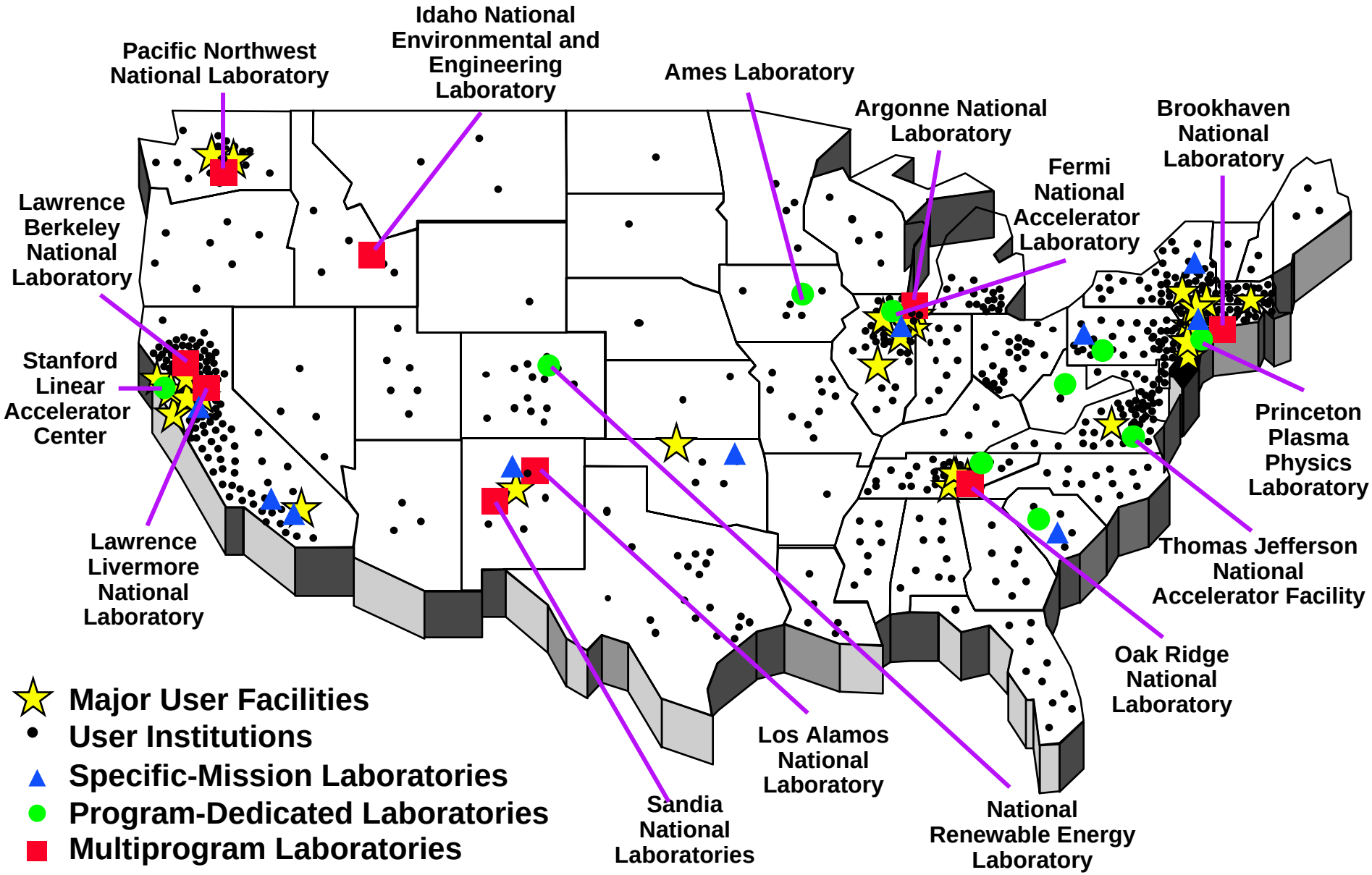
ASCAC Meeting
Washington, DC
October 17-18, 2002

Mary Anne Scott

Why is it important to consider a strategy for SC network infrastructure now?

- Growing needs for high performance networking
  - Resources are limited !
- End-to-end performance
  - Merely increasing backbone bandwidth is not the answer
- Increased need for advanced services
  - Grids are being deployed
- Advances in optical networking and rapid sweeping changes in telecommunications industry
  - Alternative business models
  - Price/performance opportunities
- Computational leadership being challenged
  - But, what about the network ?

Collaboration and Networks Critical for DOE Science



Methodology

- Engage Office of Science program offices in network planning based on a vision for future science
- Convene a workshop to bring together visionaries
  - Application representatives from high impact, high visibility initiatives
  - Network providers
  - Network and middleware researchers
- Report findings for opportunities and a path forward

- Develop strategy for Office of Science networking driven by scientific needs
- Develop a roadmap and program plan for networks

Workshop Charge

- Focus on "What is possible in the realm of science?"
  - Science unfettered by communications – scenarios
- How do the high impact applications requirements impact…
  - Network provisioning?
  - Network research?
  - Middleware research?
- Are there alternative business models that make sense in the context of the scenarios?

Workshop Findings

- Role of Advanced Infrastructure in Realizing DOE Science Visions
- Enabling Middleware Research
- Enabling Network Research
- Network Provisioning Model
- Governance Model
- Path Forward

Workshop Record: doecollaboratory.pnl.gov/meetings/hpnpw/

Advanced Infrastructure Enables DOE Science

- Much of science is already a distributed endeavor, or rapidly becoming so
  - Therefore high performance network infrastructure is critical
- There is considerable commonality in services needed to support science
  - Therefore we can define an infrastructure for distributed science
- Services must allow science to scale in many ways, by (numbers of)
  - Participants in a collaboration
  - Amount of data managed
  - Diversity in data use
  - Independent simulations, measurements, and analyses that can be combined in a single experiment
- Science paradigm shifts depend critically upon an integrated advanced infrastructure well beyond today's
  - Such shifts in scale and productivity are not speculative, as several communities are already pushing the limits
- Revolutionary shifts in how science is done can only arise from a well-integrated, widely deployed, and highly capable distributed computing and data infrastructure
  - All together, not just any one element

High Priority Middleware Research Areas

- Secure control over who does what
  - A prerequisite for any distributed science scenario
  - Challenging demands of DOE science applications and the distributed, multi-institutional nature of DOE research leads to unique demands
- Information integration and access
  - Ability to discover and access networked scientific information and information about resources
  - Enables “data-mining-based” science
- Coscheduling and quality of service
  - Ability to coordinate multiple distributed resources to provide the required performance guarantees
  - Critical to many application scenarios, including coupling of experiments with computation, data analysis pipelines, visualization and collaboration
- Effective network caching and computing
  - Stage large quantities of data to intermediate locations
  - Obtain rapid access to computing for data filtering, experiment decision-making, etc.
  - New to science Grid context
- Services to support collaborative work
  - A wide variety of “community services” designed to facilitate collaborative work, new and existing
  - Design, deployment, and operation raise challenging technical and policy issues
- Monitoring and problem diagnosis
  - End-to-end, top-to-bottom monitoring and diagnosis capabilities are essential to all distributed science applications
  - DOE researchers lead the way, but no comprehensive solution exists

High Priority Network Research

- Ubiquitous monitoring and measurement infrastructure
  - Middleware often requires understanding of the underlying network to make decisions
  - Monitoring data must be published in a scalable format, understandable by the middleware
- High-performance transport protocols
  - TCP has well-known performance limitations
  - Research into both improving TCP and into new protocols is necessary
- Multicast
  - Large, distributed collaborative projects are increasingly common in DOE, e.g., Access Grid use
  - IP-multicast is a fragile technology today
  - Mechanisms are needed to make IP multicast more robust
- Guaranteed performance and delivery
  - Need to determine what network service model will satisfy DOE research needs, and work across a variety of sites and networks
  - Trade-off between predictability and reliability
  - New approaches to network management needed
- Intrusion detection
  - Main unsolved problem is predictive analysis
  - Based on what happened in the recent past, one can get a warning an attack is about to occur
- Distributed systems vs. firewalls
  - Vetting traffic through firewalls is increasingly hard because of increased traffic, much encrypted
  - Mechanism is needed to integrate Grid security with firewalls, so the firewall can allow authorized streams

SC Needs an Integrated Network Provisioning Model

Over time, services, capabilities and app's migrate

1. Production Level Networking
   - In support of base DOE science requirements
2. Resources for High Utilization Science
   - In support of challenging science applications
   - Providing both capability networking and advanced services
3. Resources for Network Research
   - Easily separable for running controlled experiments

Challenge: Create an integrated governance model.

A Path Forward

- Analyze Requirements
  - Services the community needs for each element
    - Production Networking
    - Resources for High Utilization Science
    - Resources for Network Research
  - Opportunities providing these services
- Evaluate multiple opportunities
  - Some are time-critical, e.g., National Light Rail
- Develop an integrated roadmap
  - Across programs and between infrastructure components