US Department of Energy

Office of Science

Office of Advanced Scientific Computing Research

Division of Computational Science Research & SciDAC

High-Performance Network Research Program

Thomas D. Ndousse, PhD
Program Manger

Science Drivers

- **Data Tsunami: Terabyte – Petabyte – Exabyte**

  Emerging large-scale science experiments (LHC, ITER, SNS, climate modeling, etc.) will generate large amount of data to be distributed for analysis

- **Petascale and exascale simulation will also generate unprecedented amount of data**

- **Networks with unprecedented throughput, agility, and security will be needed to:**

  (a) Effectively and securely interconnect national and international science facilities

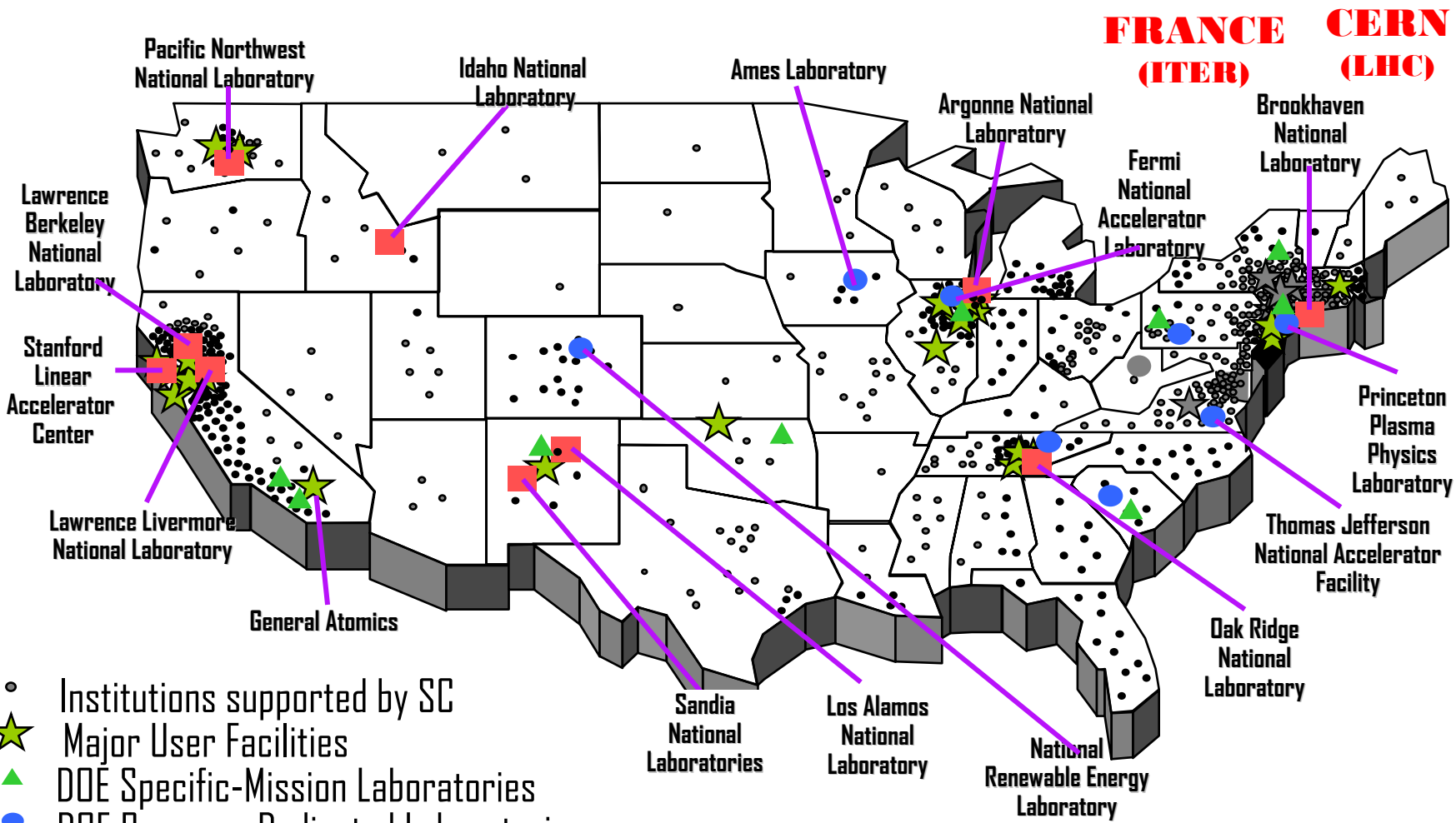
  (b) Efficiently manage and distribute massive data estimated in petascale and exascale

  (c) Tele-operation technologies for efficient use international large-scale science facilities (LHC, ITER, etc.)

U.S. Department of Energy

Office of Science

DOE/Office of Science Complex



Large-scale Distributed Scientific Collaborations

U.S. Department of Energy

Office of Science

DOE Investments in Networks for Open Science

- **ESnet - $25M/year: A high-performance production network facility supporting scientific research activities**

- **High-performance network research program ($15M/year) for developing advanced network technologies to support DOE science mission**

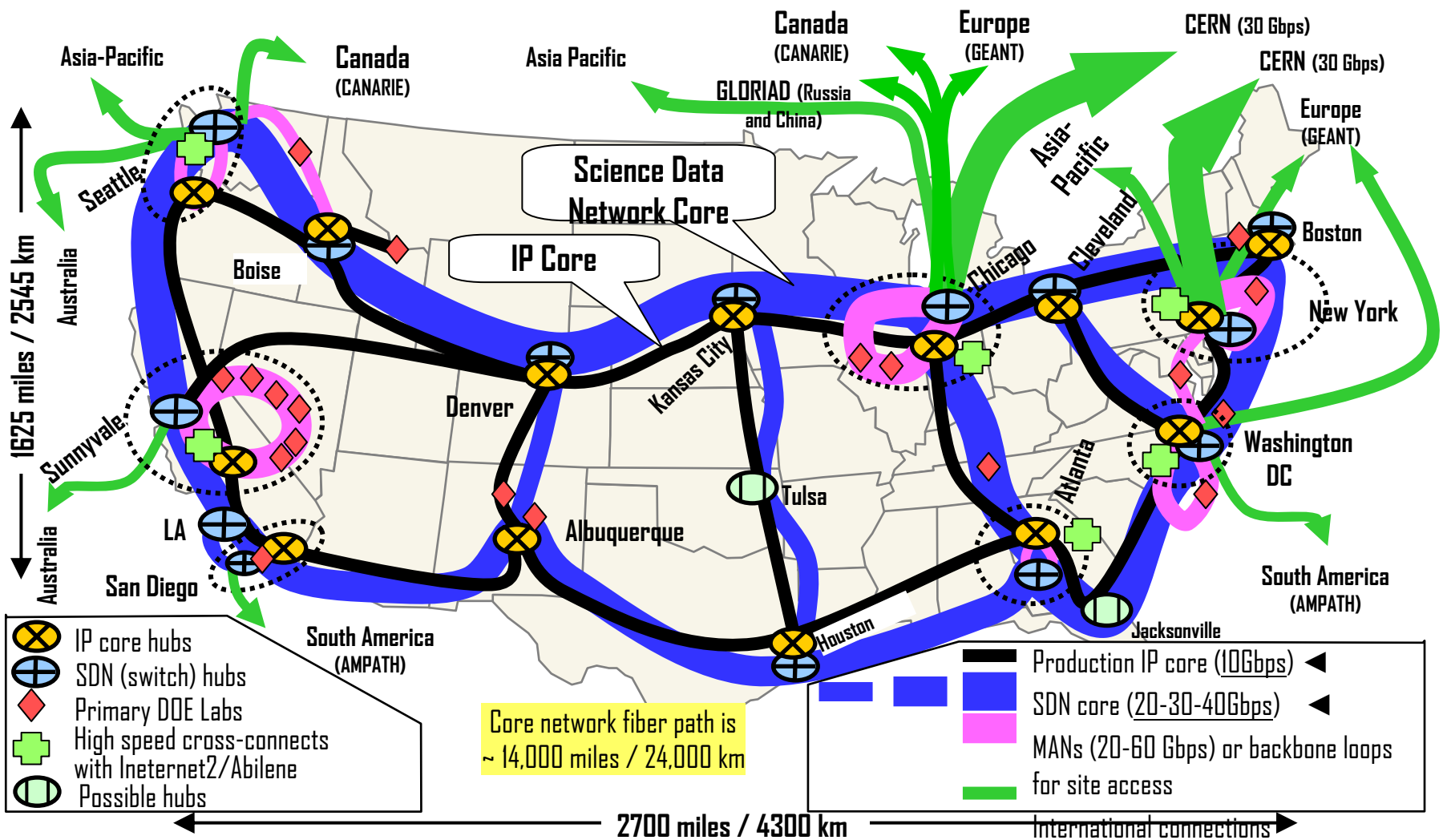
- **Experimental network facilities to prototypes advanced network capabilities that are feasible but not commercially available for production quality**

U.S. Department of Energy

Office of Science

ESnet4

60 Gbps in 2008
500 Gbps by 2010



U.S. Department of Energy

Office of Science

Network Capabilities Petascale-Exascale Science

- **Petascale/exascale computing**
  - **Leadership computing facility at ORNL (Petaflops by 2008)**
  - **Leadership computing facility at ORNL (Petaflops by 2010)**
  - **NERSC computing facility at LBNL (Petascale by 2009)**

- **Sources of Petabyte/Exabyte-scale scientific data**
  - **A) Computing at the petascale**
  - **B) Large-scale science experiments (ITER, LHC, SNS)**

- **Large-scale distributed scientific collaboration using petascale/Exascale facilities and data archives**

- **Petascale/Exascale will require Terabits Networking**

Terabits Networks Challenges

- What aspects of traditional networks components, protocols, routing algorithms, traffic approaches will be broken when networks operate terabits/sec level

- The Terabits Networking:
  - Terabits/sec in core (aggregate and single flows)
  - Multi-Gigabits/sec in hosts (single flow)
  - kbps/sec to Sensors

- The limits of packet-switching for terabits networking:
  - How well will IP/TCP, packet-based QoS, packet capture/processing (IDS), network failures, peering, etc., perform in terabits networks

U.S. Department of Energy

Office of Science

Specific Networking Issues Unique to DOE

- **Dynamic multi-capabilities networking (Differentiated capabilities Networking) to meet diverse science applications needs**
- **High-throughput transport protocols, host stacks, and e2e data distribution services (can TCP/IP do the job?)**
- **Terabits-capable cyber-security systems for open science**
- **High-performance middleware and federation networking services for distributed petascale science and large-scale scientific collaboration**

Next-Generation Transport Protocols Features

- Dynamically reconfigurable/composable and easily adaptable to different applications and different transport networks (optical, satellite, wireless, sensornets, etc.)
- Easily optimizeable when operating in shared (packet-switched), dedicated network environments (circuit-switched), and other emerging networking paradigms
- Dynamically reconfigurable to operate efficiently in terascale networks (single flow) and sensor networks
- Has the capability to implement on-demand cyber security capabilities
- Other future proof features

Dynamic multi-capabilities networking

- **Network virtualization:**
  - Each partition created dynamically on-demand to meet specific of science application
  - A partition may be at any of the network layers
  - Peering of partitions in inter-domain (federation) networking setting

- **Control Plane Technologies for Network Virtualization**
  - In-band/out-of-band control plane (centralized/decentralized)
  - Nested control planes and inter-domain control planes technologies

- **Current Network Virtualization in DOE (Hybrid Optical network)**
  - 3 partitions (layer 3 – IP network Layer networking 2.5 dynamic Ethernet circuits, and layer 2 dynamic SONET networks)
  - 4 partitions network (Partitions in layers 3, 2,5, 2 and 1)

U.S. Department of Energy

Office of Science

Experimental Networks in DOE

- **Experimental Networking facilities**
  - **Ultra-Science Network Testbed** – A 20 Gbps nation-wide layer 2 optical/SONET network for developing out-of-band centralized/decentralized control plane technologies.
  - **On-Demand Secure Circuits and Advanced Reservation Systems (OSCARS)** – An out-of-band MPLS-based control plane technology design to guaranteed multi-domain e2e services across DOE/ESnet, Internet2, and European DANTE networks
  - **Virtual Optical Networks Testbed** - A planned extended GMPLS-based control plane for layer 1 networks to be integrated to integrated with OSCARS

U.S. Department of Energy

Office of Science

DOE Ultra-Science Net Testbed

- Dynamic of layer 2.5 and 2 circuits provisioning
- Secure out-of-band control Plane
- Just-in-time scheduling and reservation
- Centralized scheduling and reservation
- Hybrid networking: integrated packet and circuit-switched capabilities
- Decentralized scheduling and reservation (GMPLS extensions)
- Cascaded control planes, nested control planes

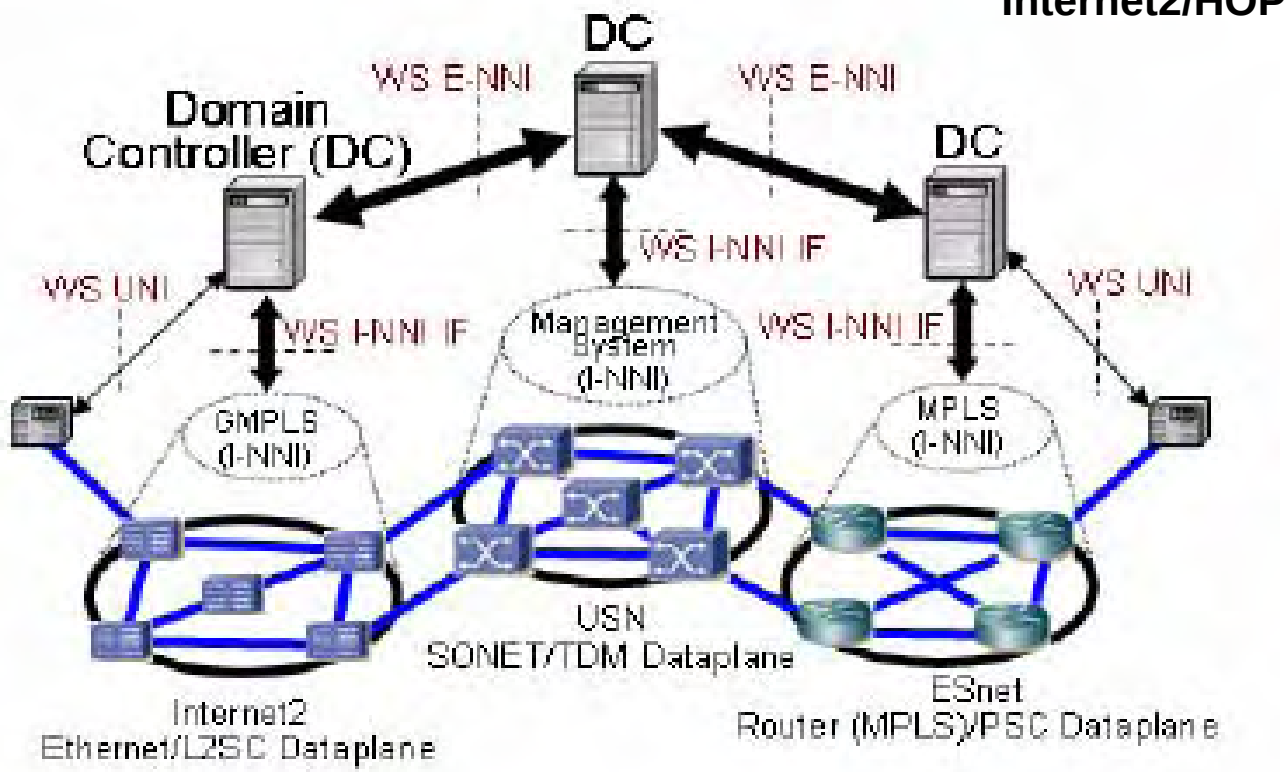
Managed by Oak Ridge National Laboratory

U.S. Department of Energy

Office of Science

On-Demand Secure Circuits and Advanced Reservations Systems (OSCARS)

- Control Plane in the IP layer
  - Data plane in layers 3, 2.5 and 2
- Participants (ESnet, Internet2/HOPI, DANTE, Dragon)



U.S. Department of Energy

Office of Science

Layer 1 Optical Virtual Networks

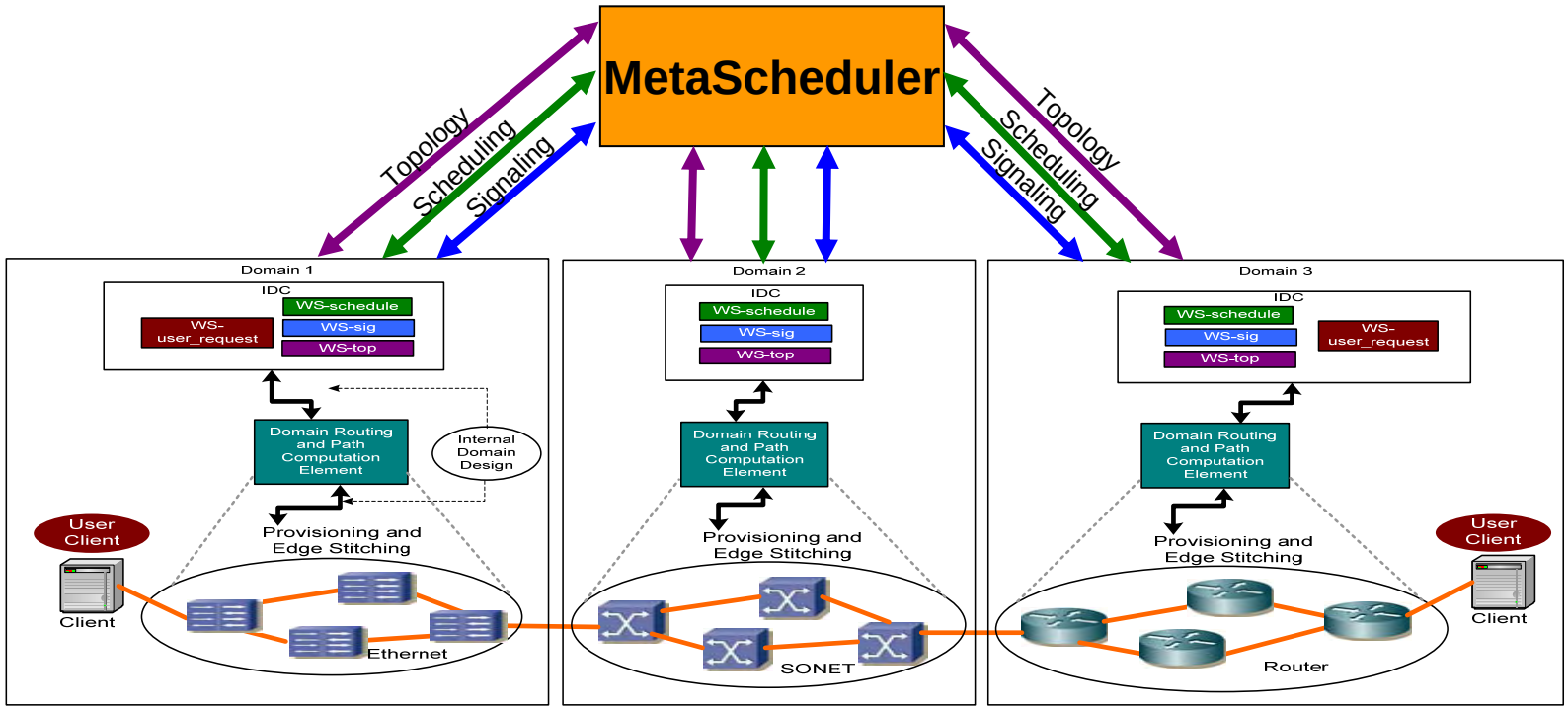
- Dynamic creation of networks in the optical layer (layer)
- Out-of-band control plane for layer 1 partition
- Joint effort of ESnet and Internet2 within Infinera/Level3 infrastructure
- Vertical and horizontal integration control plane technologies across different layers

U.S. Department of Energy

Office of Science

Hybrid Networks Control Plane

ESnet OSCARS, NSF DRAGON, and the DICE



- Meta-Scheduler Approach
- Same set of Web Services used for linear instantiation model can be used by a high level process to build services:
  - Topology Exchange, Resource Scheduling, Signaling, User Request
- A key issue is that this requires a trust relationship between the “meta-scheduler” and all the domains with which it needs to talk

U.S. Department of Energy

Office of Science

Lessons Learned from DOE Experimental networking Activities

- A robust and secure multi-domain control plane is critical to providing guaranteed e2e QoS services beyond Best-effort IP
- The core technology that will make network dynamic virtualization at each layer possible is a robust and secure control plane is
- We don't have a good understanding on how to design and build cascaded and nested control planes
- We need to extend existing routing algorithms to accommodate advanced reservations and scheduling
- Transport protocols should be dynamically optimized to the type of applications using it and to the transport networks serving the application

Related Activities

- **Mathematical Research Challenges in Optimization of Complex Systems Workshop (Dec 7-8, 2006,** Marriott Bethesda North Hotel and Conference Center)
  - **Large-scale networks proposed as a complex system**
  - **Would lead to an R&D Call**

- **Mathematics for large data sets**
  - **Under discussion for possible workshop and call**

- **Network and Cyber Security Science**
  - **New fundamental theories and mathematics that can enable the development, understanding, management, and operations of engineered complex systems like networks**