Preliminary Report & Highlights from the Subcommittee for the Examination of the Role and Efficiency of Networking and Networking Research within the Office of Science

Ellen B. Stechel, Presenting

6-November-2007

The Charge:

"weigh & review the organization, performance, expansion, and <u>effectiveness</u> of the current operations of ESnet. .. consider the <u>proposed evolution</u> of ESnet, its appropriateness and comprehensiveness in addressing the <u>data communication needs</u> .. that will <u>enable scientists nationwide to extend the frontiers of science</u>. .. make suggestions and recommendations on the <u>appropriateness and comprehensiveness of the networking research</u> .. with a view towards meeting the <u>long-term networking needs</u> .."

The Distinguished Sub-Panel

Last Name	First Name	Affiliation
Bailey	Ron	ASCAC Member
Cheswick	Bill	Lumeta Corporation (on sabbatical)
Corones	Jim	Krell Institute
Foster	Ian	ANL & U. of Chicago
Hitchcock	Dan	OASCR Liaison
Huntoon	Wendy	Pittsburgh Supercomputing Center (PSC)
Newman	Harvey	California Institute of Technology (Caltech)
Rahn	Larry	SNL & DOE-BES-Chemical Sciences
Simon	Horst	ASCAC Member
Sobieski	Jerry	Mid Atlantic Crossroads (MAX)
Stechel	Ellen	Chair & ASCAC Member
Williams	Dean	LLNL
Wing	Bill	Co-Chair & ORNL

Interpreting the Charge

- Cornerstone and Framing:
  - **"that will enable scientists nationwide to extend the frontiers of science."**
- Other Key Phrases
  - With respect to ESnet
    - ***Expansion and effectiveness of current operations, Proposed evolution***
    - In light of the Lehman review, this is de-emphasized and the panel's report will primarily site that review
  - With respect to needs
    - ***Data communication needs***
    - ***Long-term Networking needs***
      - **The sub-panel interpreted its charge broadly in considering how the network will affect all major areas of future cyber-infrastructure and its anticipated role in enabling new and transformative science processes over the coming decade.**
  - With respect to Networking Research to meet long-term needs
    - ***Appropriateness and comprehensiveness of the Networking Research***
    - The panel has been focusing more on motivating why the office should re-invigorate its networking research program
- Timeframe – **5-10 Years**
  - The network technologies and services considered will require substantial and sustained efforts to realize

Working Report Title

Implications of Network Technologies
and
Advanced Networking Research Strategies
to the
DOE Science Community

Implications of Network Technologies
and
Advanced Networking Research Strategies
to the
DOE Science Community

Timeline

- Panel Reformulated – Completed Jan 2006
- Teleconferencing, Web Meeting Capability, and Website Established – January 2007
- Two Teleconferences in February
- Status Report – ASCAC Tues 27-Feb-2007
- Bi-Weekly Meetings On-going – 13 to Date
- ESnet Visit (13-April-2007)
- Status Report – ASCAC 14-Aug-2007
- Bi-Weekly Teleconferences Continued
- DRAFT Preliminary Report–Intended Distribution to ASCAC
  - ~23-Oct-2007
- DRAFT Preliminary Report – Actual Distribution to ASCAC
  - 02-Nov-2007
- Preliminary Report – ASCAC 6-Nov-2007
- Final Report – Jan 2008

Five Baseline Observations

First: Major Drivers for Advanced “Cyber” Resources

- Computational *modeling, simulation and analysis* has grown within the e-science community to constitute a major driver for advanced cyber-resources.
- Beyond the basic computational requirements, these applications depend upon tightly integrated access to storage facilities, visualization facilities, and advanced high-performance networks.
- Being enabled by an emergence of global-scale networked systems

Second: Fundamental Changes in Data Distribution and Use

- There is an accelerating trend in the quantity of raw data collected and archived for re-use.
- In addition, the diversity and variety of data captured from instruments, computations, and sensors is growing.
- Furthermore, scientists are relying increasingly on data in its electronic format and in its derived and annotated form to pursue science in fundamentally new ways.
  - Being enabled by an emergence of global-scale networked systems

Third & Fourth: Virtual Facilities System of Systems

- The continuing trend is towards global-scale networked systems where science teams, science facilities, and the science itself crosses a wide-range of boundaries, including traditionally distinct disciplines, funding agencies, and nations.
- An increasing trend toward the development of new cyber-systems (virtual organizations, virtual facilities) by combining and integrating existing facilities in novel ways to build ever more sophisticated and more effective "e-science" environments is driving a need for a "System of Systems" approach to scientific infrastructure.
  - Will only be enabled through the emergence of global-scale networked systems

System of Systems Typically Exhibit the Behaviors of Complex Adaptive Systems

- Operational Independence of Elements
- Managerial Independence of Elements
- Evolutionary Development
- Geographical Distribution
- Inter-disciplinary
- Heterogeneity of Systems
- System of Networks
- Emergent Behavior

System of Systems Science is an emerging discipline not yet claimed by anyone, but can provide a critical organizing principle

See for example: Carlock, P.G., and R.E. Fenton. "System-of-Systems (SoS) Enterprise Systems for Information-Intensive Organizations," Systems Engineering, Vol. 4, No. 4 (2001), pp. 242-261.

The Office of Science facilities infrastructure represent a System of Systems

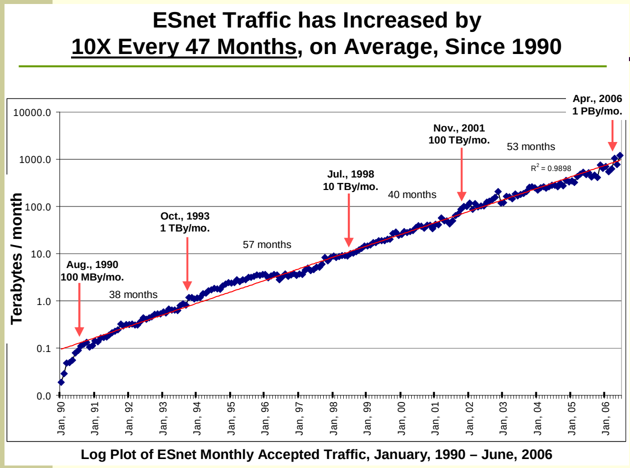
- **Complex Systems with Many Similar and Different Components**
  - Producing data at a phenomenal rate
  - Producing large amounts of heterogeneous, geographically dispersed data
  - Relationship and coupling of data and computation
  - Cultural barriers
  - Growing in its dependence on the emergence of advanced networks
- **Systems, Projects, People**
  - Computing System
  - Experimental Facilities, Instruments, and User Centers
  - Observational Networks
  - Data, Information Resources, and Visualization
  - Virtual Organizations and Distributed Collaborations
  - OASCR Research Projects:
    - Applied Math, Computer Science, Networking Research, SciDAC
  - Researchers, Facility Staff, Center Staff, Program Managers
- **The Emergent Behavior is Breakthrough Advances in Scientific and Engineering Research**
  - Accelerating and amplifying impact on urgent problems of global scale

Fifth: Emerging Network Technologies Transformative in Character

- The network now underlies nearly every aspect of advanced distributed science activities.
- Recent advances in network technologies could provide unprecedented increases in global reach, capacity, performance, and user manageability and control.
- Hybrid (packet + circuit) networks are an example of one such emerging network technology that is ***transformative***
  - Re-define how one can conceive and construct large-scale global networked environments to advance science in fundamentally new ways.

Seven “High Level” Findings

First: About ESnet



- ESnet facility has done an exemplary job in collecting, analyzing, architecting, deploying, and operating a very high performance and effective network infrastructure to serve the DOE science needs.
- This facility has successfully, to date, been based upon a continuous “requirements” driven process.
- The panel feels that the ESnet4 infrastructure is critical in the relative near term, and will play a continuing vital role in the long term as new network architectures and services evolve and mature.
- Sustained and adequate funding for these facilities remains essential

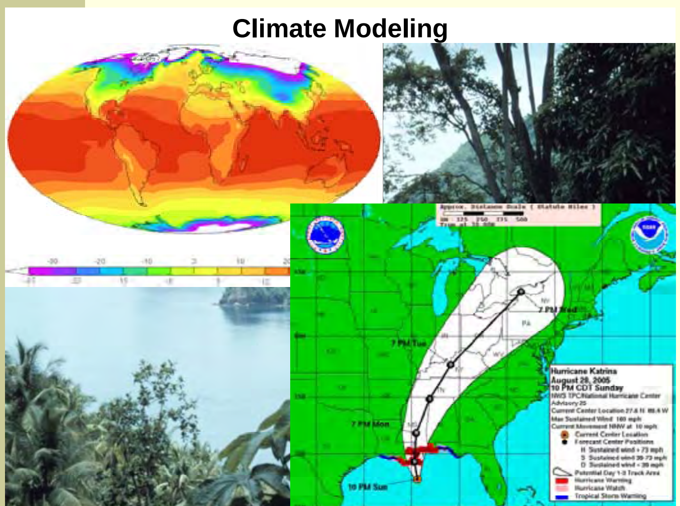
Second: DOE Science Cannot Depend Solely on the Commercial Sector

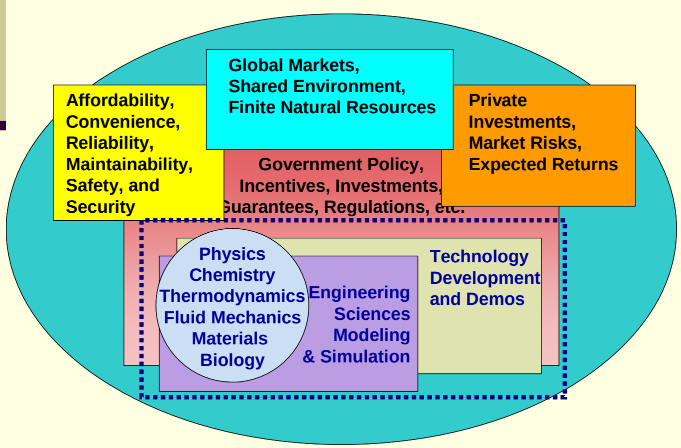
- In the coming decade, much of the network technologies necessary to meet DOE (and other science-based agency) e-science requirements will not arise naturally from commercial R&D.
  - Commercial R&D is focused on e-commerce and the Internet,
    - Needs orthogonal to & independent of the needs of e-science.
  - DOE is at the forefront of developing peta-scale science
  - DOE must lead other agencies in the development of a Peta-scale network and networking services.
- The ***challenge*** is to construct a long term and far-reaching network research and development program that sustains technology innovation from basic research to prototype to early deployment and ultimately culminates in the production network facilities.

Third: Enable Systems Level Science

- Networks have played an ever-increasing role in the pursuit of DOE science over the last thirty years.
  - The role of the network in future science will accelerate this trend.
  - Uniting distributed teams,
  - Allowing the efficient exchange of information among organizations and science applications, and
  - Enabling new cyber-infrastructure capabilities and
  - Ultimately must enable system-level science.
- The ***challenge*** is to develop network architectures and service models that support such virtual organizations, and provide reliable, predictable, and repeatable network performance, accessibility, and security.
  - Quantified service
  - End-to-end service (application to application)
  - Work-flow provisioning and management, and
  - Federated trust.

Systems-Level Science - Definition





- The integration of diverse sources of data, information, and knowledge about the component parts of a complex system
- ***Goal*** to obtain an fundamental and *predictive/descriptive* understanding of the whole system's properties.
- Such system-level approaches are becoming increasingly important in a number of fields
  - Clean & Secure Energy
  - Climate Change
  - Human Health
  - Unconventional Warfare

Fourth: Profound Impact from National and Global Data-Intensive Distributed "Cyber-Environments"

- A key technology that will have profound impact on society and the economy in the coming decade will be the development of national and global scale, data-intensive (terabyte to petabyte scale, and beyond) distributed "cyber-environments."
  - Such cyber-environments will be constructed from a combination of large-scale distributed computational systems, large-scale integrated data repositories, sensor networks, and scientific instruments.
  - By their very nature, such environments will rely on networks with unprecedented global reach, capacity, performance, and user control.
- The ***challenge*** is to develop dynamic and intelligent [cyber] resource allocation architectures and agents that will allow the end-user (the DOE scientist) to create and incrementally refine the cyber-environment they need for their science.

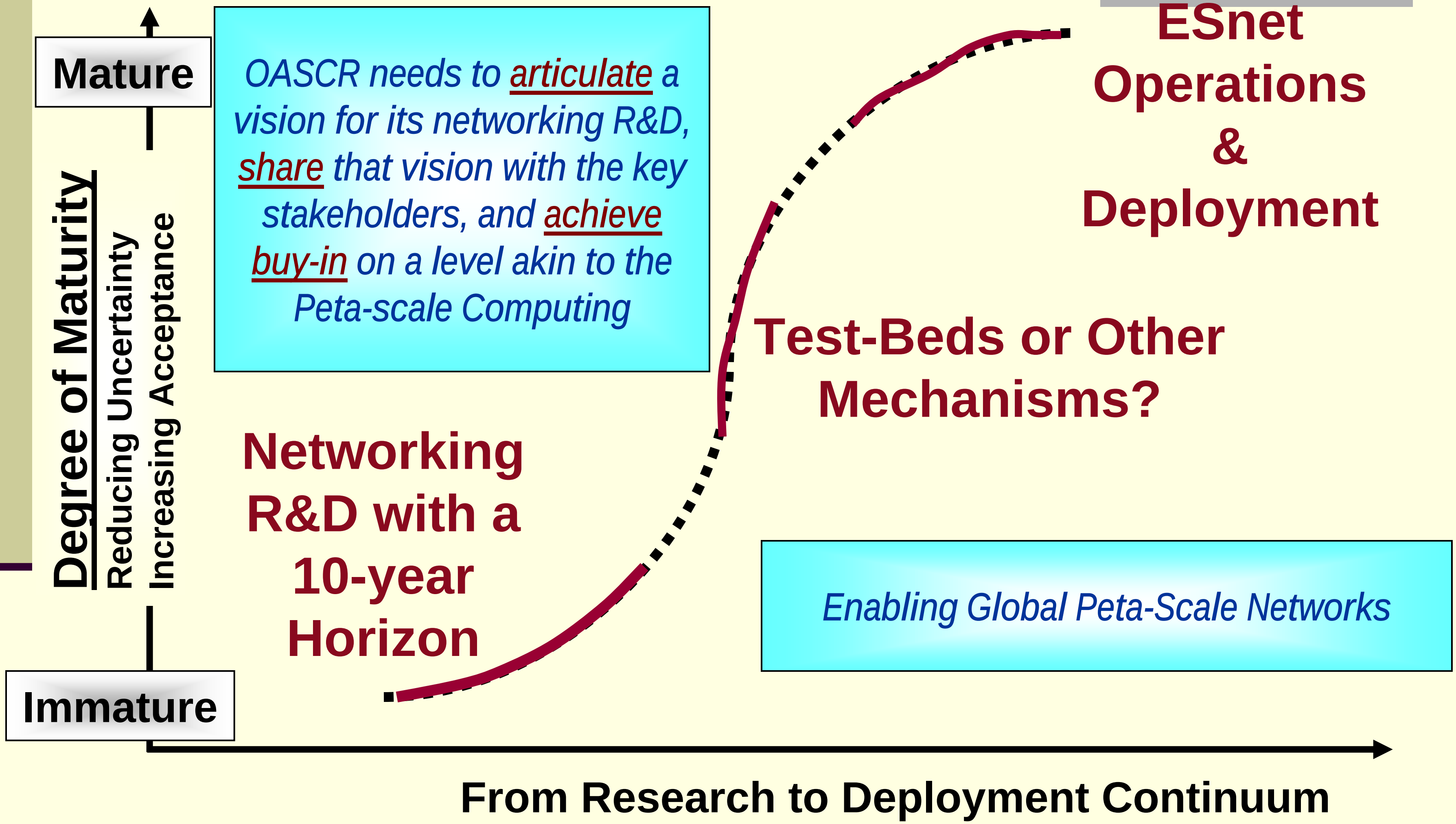
Fifth: Petabit/Second in 10 Years?

- Advanced networks are already, in aggregate, on the 100 Gigabits/second scale, and will be at the Terabit/second scale within the next five to seven years.
  - The emergence of high capacity “dense wave division multiplexing” (DWDM) optical telecommunications systems and their increasing affordability have allowed advanced applications to explore these new capabilities.
  - Existing technology development of 100 Gigabit/sec transmission expected deployable in the 2011-12 timeframe,
    - May be inadequate to meet the needs of data intensive systems envisioned.
  - The network capacity and service capabilities anticipated for future science activities will likely be three to six orders of magnitude greater than current network architectures, technologies, or common practice can effectively address today.
- The ***challenge*** is to accelerate the ten-year technology trajectory to achieve the petabit per second application in 10 years.

Sixth: Avoiding the "Valley of Death"

- Promising new network research (across all agencies and industry) often sits on the shelf and do not mature into viable new network services.
  - Sometimes (if a private sector market emerges,) such technology may be incorporated into new products or services.
  - In the DOE environment, long- term delays in finding and leveraging useful new network technologies hinder progress.
    - Our science depends critically on advances in science infrastructure
  - This limbo period between the research and the [hopefully] ultimate revival and productization is euphemistically referred to as the "***Valley of Death***."
- The ***challenge*** is to establish mechanism(s) to identify promising network research concepts and move them progressively through prototyping, experimental deployments, and ultimately into a production network service environment.

A Strategy for Moving up the Maturity Continuum and Spanning the “Valley of Death”



Seven: Capitalizing on Domain Specific Advances

- By necessity, large science programs have engaged in leading-edge development efforts pushing the state of the art in networks and networking services.
  - This in turn has led to a vast chasm between the current level of data distribution and management capabilities and that which the rest of the scientific community is able to use.
- The ***challenge*** is two fold:
  - first, is to leverage the advances of such efforts for the broader good of the DOE science community, and
  - second, is to develop advanced cyber-infrastructure service architectures that allow future programs to, easily and effectively, create the types of network (or other cyber infrastructure) that they need.

Six “High Level” Recommendations

Recommendation #1

- The Office of Science (SC) should recognize the **network and networking services as an explicit and fundamental element of advanced peta- and exa-scale science,** of equal importance to leadership class computing and large scale scientific facilities in **enabling high-end science.**
- The sub-panel ***recommends*** that the department approach the development of advanced networking in a fashion similar to the manner in which it develops its goals and objectives for peta-scale and exa-scale computing.
- Hence, the sub-panel encourages the office to create mechanisms whereby the DOE science community is encouraged to think broadly about how an unconstrained network resource might transform the conduct of science.

Recommendation #2

- SC should recognize that their **scientific facilities and science programs operate with the characteristics of a "System of Systems."**
- These facilities and programs, with their specializations and focus, will increasingly be required to interoperate and inter-communicate with one another to realize multi-disciplinary or systems-level science.
- Advanced information technologies will enable these Systems of Systems, and advanced networking architectures and capabilities will be particularly critical.
- The ***recommendation*** is that OASCR should establish mechanisms to investigate the implications and issues that arise from the concept of a System of Systems, specifically as the concept applies to DOE science, facilities, and programs.

Recommendation #3

- An **aggressive, sustainable, long term, and strategically focused network research and development program** must be **re-invigorated** to create network specific technologies that will allow DOE to transform the manner in which science is done - not just increase the speed of existing systems.
- The ***recommendation*** is that OASCR should convene an external committee to review this networking research program on a regular basis, in order to, continually, maintain a ten-year research horizon.

Recommendation #4

- Any **network research program** must have a **strategy to bridge the "valley of death"** and have a **deployment objective** beyond publishing a scholarly paper (if successful.)
- Hence, OASCR needs to develop a deliberate strategy for moving concepts (whether from within their basic network research program or from the network research community more generally) through test-bed like deployment and finally through to production.
- The "*valley of death*" should be addressed with funding to support applied R&D and experimental/early adopter deployments.
- Furthermore, the process should also involve increasing collaborations between the network research scientists and operational/engineering facilities personnel as concepts mature.
- Moreover, the end-user, application science community must be a continual and integral component of the collaboration.

Recommendation #5

- It must be recognized that the effective use of **next generation networks**, with unprecedented and rapidly expanding capacity and utility, as well as complexity, **will require a new paradigm of operations and management**, including:
  - end-to-end monitoring of the network (necessarily including the end systems,)
  - use of autonomous software agents to operate, manage, diagnose, and alert as needed, and
  - use of higher level services to ensure optimal network resource utilization and workflow coordination and management.
- The ***recommendation*** is that priority research directions within the network research program should include
  - the exploration and development of automated, dynamic, intelligent agents to take full advantage of emerging network architectures, with explicit goals of furthering DOE Science.
- OASCR should also convene a workshop to flush out additional priority research directions.

Recommendation #6

- Advances in data collection, archiving, curation, generation, pedigree, and access are essential in the emerging cross-disciplinary distributed science environment and for systems-level science.
- Effective integration of new networking technologies into the future data-management architecture is crucial to providing timely and secure access and/or efficient migration of these large data sets across a global user base.
- The sub-panel ***recommends*** establishing a data management (and all its dimensions) component of the network research program in order to integrate these two symbiotic aspects of information technology.

Questions/Discussion

Thank-you for your Attention