Campus Cyberinfrastructure (CC*DNI) and Performance Measurement

Kevin Thompson
NSF CISE/ACI
January 2015
This talk has 3 parts

- A few words on NSF
- A few words on the new solicitation CC*DNI
- Some thoughts on performance measurement
NSF Core Mission: Fundamental Research

- $7.1 billion FY 2014 research budget
- 94% funds research, education and related activities
- 50,000 proposals
- 11,000 awards funded
- 2,000 NSF-funded institutions
- 300,000 NSF-supported researchers

Fundamental Research
Computer and Information Science and Engineering (CISE) Directorate

*Exploring the frontiers of computing*

- Promote progress of computer and information science and engineering research and education, and advance the development and use of cyberinfrastructure.
- Promote understanding of the principles and uses of advanced computer, communications, and information systems in support of societal priorities.
- Contribute to universal, transparent and affordable participation in a knowledge-based society.

*These frontiers have interfaces with all the sciences, engineering, education and humanities and a strong emphasis on innovation for society.*
NSF Cyberinfrastructure (ACI) is part of the CISE Directorate and responsible for NSF-wide CI coordination and support.
ACI Mission: support advanced cyberinfrastructure to *accelerate* discovery and innovation *across all disciplines*

- Coordination role across NSF
- Interagency & international partnerships
- Supports Use-inspired Cyberinfrastructure for research, education
- Inherently multidisciplinary with strong ties to all disciplines/directorates
Advanced Cyberinfrastructure

Supports the research, development, acquisition, and provisioning of state-of-the-art CI resources, tools, and services

**HIGH PERFORMANCE COMPUTING**
- Enable scientific discovery via petascale computing: diverse, integrated, state-of-the-art HPC assets

**DATA**
- Support scientific communities in use, sharing, archiving of data by creating building blocks for data infrastructure.

**NETWORKING AND CYBERSECURITY**
- Invest in campus network improvements/re-engineering to support modern computational science. Transition cybersecurity research to practice.

**PEOPLE**
- Transform innovations in research and education into fully integrated, sustained software resources
NSF embraces an expansive view of Cyberinfrastructure motivated by research priorities and the scientific process

**Organizations**
- Universities, schools
- Government labs, agencies
- Research and Medical Centers
- Libraries, Museums
- Virtual Organizations
- Communities

**Expertise**
- Research and Scholarship
- Education
- Learning and Workforce Development
- Interoperability and operations
- Cyberscience

**Scientific Instruments**
- Large Facilities, MREFCs, telescopes
- Colliders, shake Tables
- Sensor Arrays
- Ocean, environment, weather, buildings, climate, etc

**Computational Resources**
- Supercomputers
- Clouds, Grids, Clusters
- Visualization
- Compute services
- Data Centers

**Data**
- Databases, Data repositories
- Collections and Libraries
- Data Access; storage, navigation, management, mining tools, curation, privacy

**Software**
- Applications, middleware
- Software development and support
- Cybersecurity: access, authorization, authentication

**Networking**
- Campus, national, international networks
- Research and experimental networks
- End-to-end throughput
- Cybersecurity
NSF Opportunities for Networking and Communication R&D

IRNC: International Research Network Connections

CPS: Cyber-Physical Systems

CRI: Computing Research Infrastructure

cyberSEES

CIF: Comm & Information Foundations

EARS: Enhancing Access to the Radio Spectrum

NeTs: Networking Technology and Systems

Campus Cyber-infrastructure (CC-NIE, CC*IIE, CC*DNI)

Smart & Connected Health

Networking & Communication R&D
Partnerships: Many dimensions

- Promoting partnerships: NSF's strategic goal
- Partnerships build capacity, leverage resources, and increase the speed of translation from discovery to innovation.
- Collaboration and partnerships between disciplines and institutions and among academe, industry, and government enable the movement of people, ideas, and tools throughout the public and private sectors.
- CISE collaborations engage partners from:
  - Other disciplines
  - Industry
  - government agencies: federal, state, local
  - International funding agencies
Get Involved

• Volunteer to be a reviewer.
• Visit NSF, get to know your program(s) and program director(s).
• Develop transformational ideas and send your best ideas to NSF.
• Participate in NSF-funded and hosted activities (e.g., workshops, COVs, ACs).
• Participate in the CCC/CRA visioning activities.
• Develop transitional ideas for how to move from ideas and prototypes to systems deployed on testbeds to technology transfer.
• Work within your institution to support and reward interdisciplinary research.
• Work within your institution to support service to the larger computing community around the globe.
• Send us your accomplishments; advertise your research to other citizens through local radio or TV, blogs, newspaper articles, etc.
• Join NSF to serve as program officers or division directors.
CC*DNI: Campus Cyberinfrastructure - Data, Networking, and Innovation

Kevin Thompson
CISE ACI
1st a quick reminder - Global topology r&e networking (NSF supports a part of this)
Zoom (note the different scales: international, national, regional, campus/local)
CC*DNI Synopsis

- CC*DNI integrates campus-level data & networking infrastructure for higher levels of performance, reliability & predictability for science applications & distributed research, with an explicit element supporting models for potential future national scale network-aware data-focused cyberinfrastructure attributes, approaches & capabilities.

- CC*DNI combines CC*IIE and DIBBS (Data Infrastructure Building Blocks) for 2015
Campus Cyberinfrastructure – Data, Networking, and Innovation (CC*DNI) Program

- FY15 new solicitation, NSF 15-534
- 7 categories of proposals, 2 of them are New
- Data Infrastructure Building Blocks (DIBBs) - Multi-Campus/Multi-Institution Model Implementations
  - Models for potential future national scale network-aware data-focused cyberinfrastructure attributes, approaches, and capabilities – sharing data beyond a single institution

- Network-centric categories (from CC*IIE)
  - Data Driven Networking Infrastructure for the Campus and Researcher
  - Network Design and Implementation for Small Institutions awards
  - Network Integration and Applied Innovation
  - Campus CI Engineer
  - Regional Coordination and Partnership in Advanced Networking
  - Instrument Networking
2015 CC*DNI Program Areas

- Data Infrastructure Building Blocks (DIBBs) - Multi-Campus/Multi-Institution Model Implementations
  - Up to $5,000,000 for up to 5 years
- Data Driven Networking Infrastructure for the Campus and Researcher
  - Up to $500,000 for up to 2 years
- Network Design and Implementation for Small Institutions
  - Up to $350,000 for up to 2 years
- Network Integration and Applied Innovation
  - Up to $1,000,000 for up to 2 years
- Campus CI Engineer
  - Up to $400,000 for up to 2 years
- Regional Coordination and Partnership in Advanced Networking
  - Up to $150,000 for up to 2 years
- Instrument Networking
  - Up to $400,000 for up to 2 years

please refer to the solicitation for complete detail
CC*DNI Program-wide Criteria

- Science Drivers
- Partnerships between campus CI experts and scientists
- Campus Cyberinfrastructure plan
  - plan within which the proposed network infrastructure improvements are conceived, designed, and implemented in the context of a coherent campus-wide strategy and approach to CI
  - Maximum 5-pg supplementary document addressing:
    - Sustainability of proposed work in terms of ongoing operational and engineering costs
    - Refer to solicitation for guidance on technical areas to include such as IPv6, InCommon federation, and IP spoofing
  - See example CI plans from existing awardees at http://fasterdata.es.net/campusCIplanning/
Regional Coordination awards made in 2014 (CC*IIE)

- #1440642 (Meehl, UCAR)/#1440568 (Hauser, Colorado)
  - “Collaborative Research: CC*IIE Region: Rocky Mountain Cyberinfrastructure Mentoring and Outreach Alliance (RMCMOA)”

- #1440774 (Monaco, Kansas St)
  - “CC*IIE Region: Leveraging Partnerships Across the Great Plains to Build Advanced Networking and CI Expertise”

- #1440659 (von Oehsen, Clemson)
  - “CC*IIE Region: Southern Partnership in Advanced Networking (SPAN)”

- #1440450 (Schopis, OSU)
  - “CC*IIE Region: Transforming a Regional Network and the Regional Community to Serve Diverse and Emerging Research Needs”

- [1 more award still pending]

- The Quilt consortium and their meetings may provide an excellent forum to share ideas in this space
Example Accomplishments from ACI Networking Programs

- (CC-NIE) UMD – developing network embedded storage and compute resources via Software Defined Networking (SDN) and exposing services to scientific applications and workflows
- (CC-NIE) U of Washington – campus networking upgrades doubled particle physics data transfers to/from PNNL to 1.4Gbps single flow (Ed Lazowska, PI)
- (IRNC) 4X capacity improvement (80Gbps aggregate) in connecting Astronomy facilities in Hawaii to US mainland
- (IRNC) 4X capacity improvement (40 Gbps aggregate) between US and South America – LSST may require 100Gbps by 2020
Another Example – U of Dayton

● Impact at Dayton – “a high performance connection...driven by our NSF strategy of providing DMZ connections for researchers with a specific need. NSF is truly helping the University of Dayton ‘raise the entire harbor’ for science and engineering work on campus and we have used the prestige of this grant to get the attention of our campus leadership to ensure the continued funding for HPC investments.

● Impact on Dayton Partners - “Our work bringing up the connections at Central State Univ (Historically Black College) has gone well. Part of our funding supports upgrades at their campus. CSU does not have deep-expertise on networking at their campus, so we sent our engineering staff to supervise the work of contractors in upgrading their fiber and connecting the new DMZ infrastructure...the NSF support has truly transformed a chunk of the CSU network into a science-ready environment that has our researchers working collaboratively with their faculty and students. The funding from NSF is making a huge difference for several faculty and students at Central State - They are working on cutting edge projects with Vijay Asari on our campus in the area of "Computer Vision”

Thomas Skill, PI (CIO) University of Dayton
CC*IIE Award Map 2012-2014

2012-2014 National Science Foundation CC-NIE & CC-IIE Awardees
Performance Measurement
PerfSonar and CC*DNI

“Proposals are expected to describe an approach to end-to-end network performance measurement based on the perfSonar framework with associated tool installation and use; proposals may describe an alternative approach to perfSonar with sufficient justification. Proposers are encouraged to reference the following community web site for more information on perfSonar: http://fasterdata.es.net/performance-testing/perfsonar/.”

#1440667 – Martin Swan, Indiana U – “CC*IIE Integration: Development and Integration of perfSONAR for End-to End Network Cyberinfrastructure”
R&D Findings and Recommendations

- Finding-5) The initial motivation to use perfSONAR for many was to provide network measurement for network engineers i.e., as a different approach to monitoring than what was traditionally used on a campus or network infrastructure. Over time this focus has shifted more toward end-to-end use cases, and use by end-users, instead of just operations staff using to monitor the network core.

- Recommendation: More capabilities that are useful to end-users (and not just operators) should be developed in perfSONAR extension development efforts in order to meet the shift in use cases compared to the use cases targeted in the original design.

- Finding-6) New operational architectures that are emerging with Science DMZ and software-defined networking (SDN) adoption have caused complications for tools like perfSONAR, which must change to provide adequate views into the changing network and user behaviors.

- Recommendation: To meet the needs of new operational use cases of perfSONAR, configurability of perfSONAR when used in Science DMZ and SDN environments must be deeply explored. Particularly, perfSONAR must be standardized to jointly function with SDN in cases where there is a potential to discover change in network topologies and to dynamically alter testing behavior or application resource provisioning.

- Finding-7) There are on-going concerns in the community regarding the risks of deploying perfSONAR software (e.g., will it pass integrity scans? or does it make my network more vulnerable for cyber attacks?).

- Recommendation: The perfSONAR community should produce a security business continuity-planning (BCP) document that describes risks and mitigations for the data, services, and hosts; perfSONAR development team should consider keeping this document in a public location and seek on-going community inputs.

- Finding-8) There is a steep learning curve to using perfSONAR effectively, which has been an adoption challenge. Interpretation of the measurement data, to actually find and evaluate network performance, still remains a challenge without a knowledgeable engineer available.

- Recommendation: Training efforts for operators and end-users are vital to foster continued adoption and long-term effective use of perfSONAR. Such efforts should be supported by funding agencies through programs that encourage campuses to create a ‘Cyberinfrastructure Champion’ to support larger-scale of perfSONAR adoption; the expertise also enables a location-specific support structure to use measurement utilities more effectively for data movement and other high-performance networking needs.
Communities are just now starting to adopt PerfSonar in distributed science environments
The larger umbrella of Performance/Network Measurement, Active and Passive

- There’s a lot going on out there, only some of it directly impacts operational R&E networks, here are some passive techniques
  - Web100/Web10G: TCP instrumentation at source/sink
  - Packet header capture / port mirroring, etc.
    - e.g. Bro, although Bro as open source IDS is not generally used for this purpose
    - Allows complete flow behavior (re)construction and profiling
    - Sampling doesn’t count in my opinion if you’re serious about performance analysis, but it's better than nothing
  - SNMP – 20th century granularity and accuracy (seconds, seriously?)
  - RouteViews – current and historical control plane state

- NSF support for Web100/Web10G, RouteViews, Bro and other R&D in net measurement – e.g. [www.caida.org](http://www.caida.org)

- And yet, in 2015, the NSF community knows surprisingly little about the nature and performance of scientific data flows.
  - Campuses know surprisingly little based on 3 years of campus network proposals and reports
  - SNMP-based interface utilization graphs are not nearly enough, campuses need to up their game to best serve their research and education mission
  - My.es.net – major step in the right direction (ESNet - DOE lab focused)
The Future?

- PerfSonar – first order as a data sharing framework
- Combining data on “what’s possible” with “ground truth” on real scientific data flows
  - e2e performance: tested vs. actual, and why
  - Integration with RouteViews, Web10G, NetFlow, SNMP
  - Passive monitoring triggering active monitoring
  - Manual vs. automated and agent-based
- Informing the NOC(s), human-in-the-loop
- Multi-AS paths even for domestic R&E traffic
  - Federated data sharing
  - International paths and inter-federated data sharing
- Ubiquity through sensor deployment
- Network measurement in an SDN world
- One challenge – will PerfSonar lead directly to documented improved scientific data flow performance end-to-end?
Credits

- Copyrighted material used under Fair Use. If you are the copyright holder and believe your material has been used unfairly, or if you have any suggestions, feedback, or support, please contact: ciseitsupport@nsf.gov.

- Except where otherwise indicated, permission is granted to copy, distribute, and/or modify all images in this document under the terms of the GNU Free Documentation license, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation license” at http://commons.wikimedia.org/wiki/Commons:GNU_Free_Documentation_License.

- The inclusion of a logo does not express or imply the endorsement by NSF of the entities' products, services, or enterprises.