# *Cybersecurity for HPC Systems: Challenges and Opportunities*

Sean Peisert
Lawrence Berkeley National Laboratory

NSCI HPC Security Workshop — Sept. 29, 2016

U.S. DEPARTMENT OF ENERGY

UNIVERSITY OF CALIFORNIA

BERKELEY LAB
Lawrence Berkeley National Laboratory

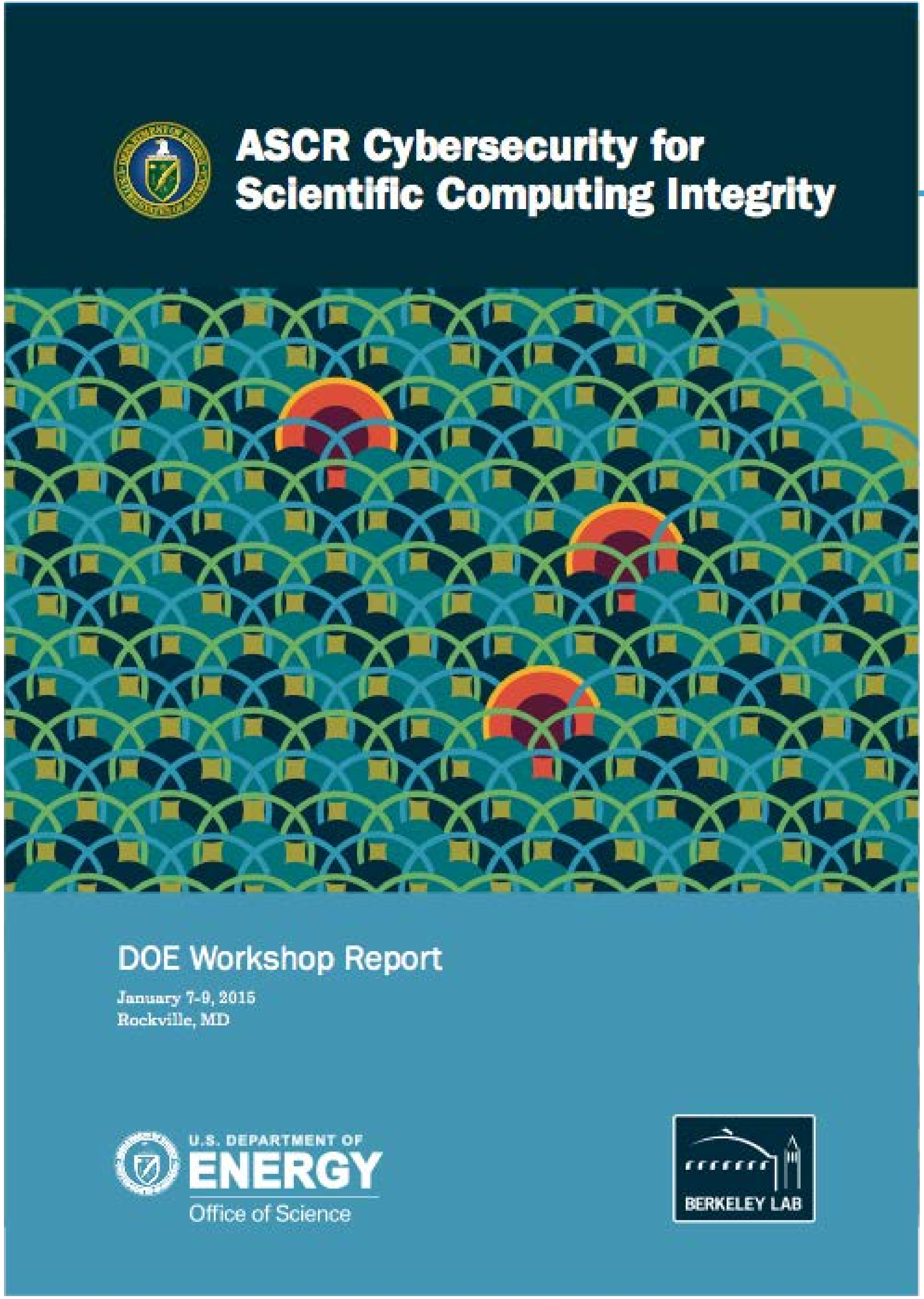
# High-Performance Computing Has Become Essential to U.S. National Security and Prosperity

- *Scientific understanding*
  - evolution of the universe
  - climate change
  - biological systems
  - renewable energy
  - aerodynamics
  - precision medicine
  - nuclear stockpile safety
- *Engineering analysis*
  - Aerodynamics/hydrodynamics
  - Materials
- *Cryptanalysis*
- ..and more

# HPC Centers are Targets

# What are the threats to HPC?

- Confidentiality
  - Data leakage (even in “open science”)
- Integrity
  - Alteration of code or data
  - Misuse of computing cycles
- Availability
  - Disruption/denial of service against HPC systems or networks that connect them



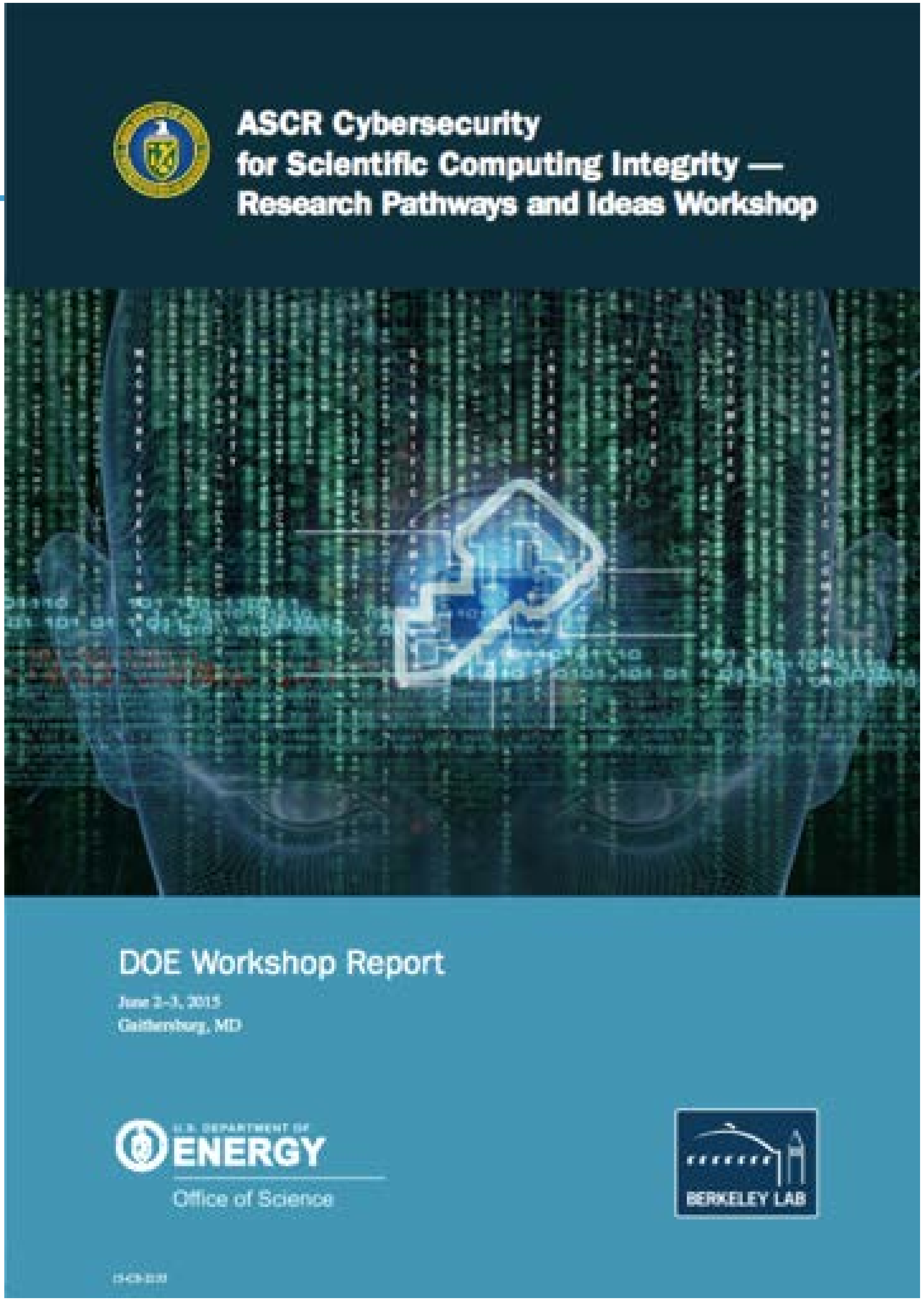
## ASCR Cybersecurity for Scientific Computing Integrity

DOE Workshop Report

January 7-9, 2015
Rockville, MD

U.S. DEPARTMENT OF ENERGY
Office of Science

BERKELEY LAB



## ASCR Cybersecurity for Scientific Computing Integrity — Research Pathways and Ideas Workshop

DOE Workshop Report

June 2-3, 2015
Gaithersburg, MD

U.S. DEPARTMENT OF ENERGY
Office of Science

BERKELEY LAB

[illegible]

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

# Insiders are Important to Consider

- ***Insider*** — someone who has some combination of:
  - ***access*** to a resource,
  - ***knowledge*** of an organization, and/or
  - ***trust*** by an organization.

  - There can be degrees of this.

- ***Insider threat*** — threat posed that an insider ***may*** abuse their discretion
  - Can also be degrees of this.

- ***Insider attack*** — action or attempted action with the potential to violate the system's security policy.

# HPC and Traditional IT: Similarities

- Similarities
    - On the surface…
        - Connected to IP networks
        - Often Linux-like OS
            - Similar hardware, software, & configuration flaws as other systems

# What makes security for HPC different?

- HPC systems tend to:
  - have very *distinctive modes of operation*; or
  - be *used for very distinctive purposes*, notably mathematical computations;
- Some HPC systems:
  - run highly *exotic hardware and software stacks*, and/or
  - are *extremely "open"* to users.

- *This distinctiveness presents both **opportunities** and **challenges***

# HPC Security Challenges

# Constraints to security in HPC environments

- High performance!
  - Computation
  - Data transfers

- Also, many HPC systems (NSF, DOE ASCR) are extremely open, including international collaborations.
  - Can’t just “air gap” the HPC system.

∴ Can’t use certain security solutions, such as network firewalls in the same way

# How do IT security solutions work for HPC?

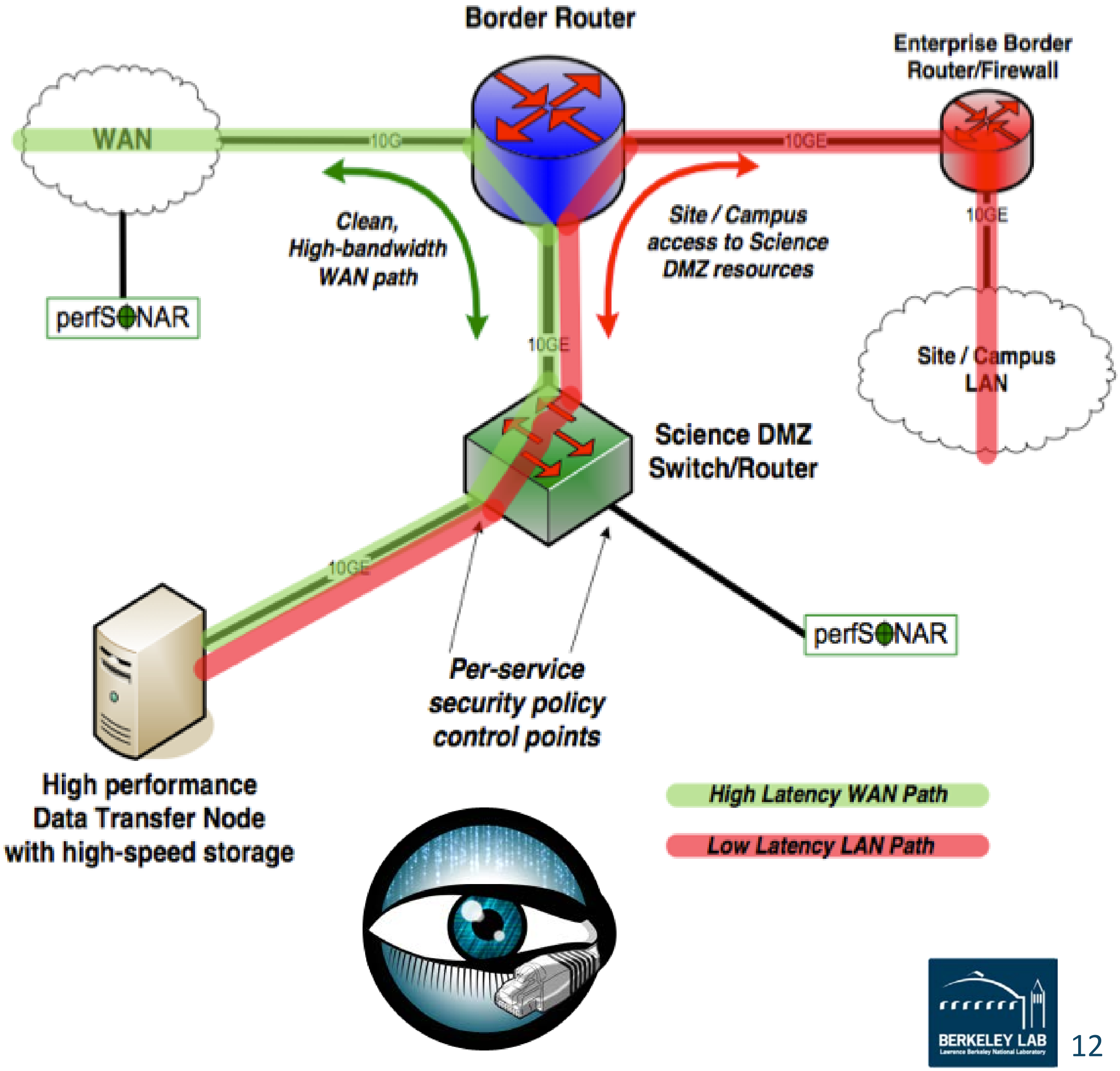
- Traditional IT security solutions:
  - network and host-based intrusion detection
  - access controls
  - (maybe) lightweight software verification

- … they work about as well in HPC as traditional IT (often not very), or worse, due to constraints in HPC environments.

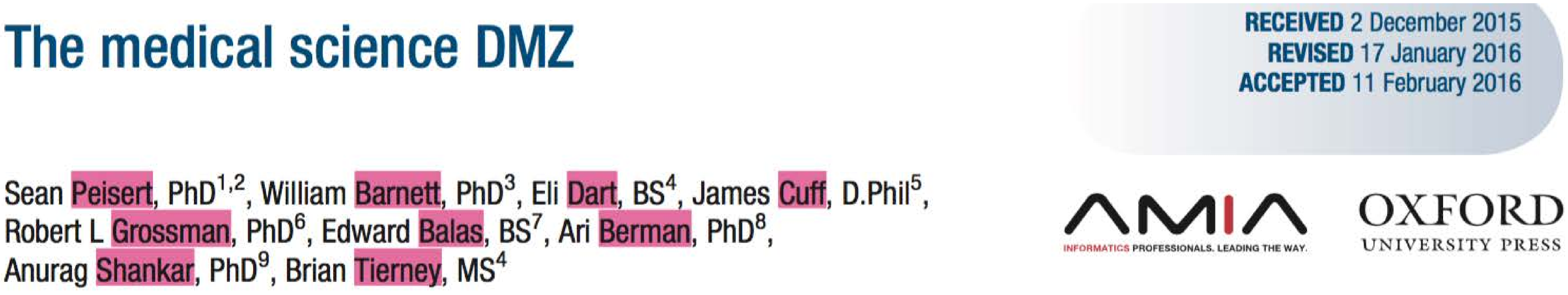
- *Some solutions exist that can help compensate for these constraints*

# Science DMZ

- Security model that optimizes network throughput
  - Isolates a site’s scientific computing in its own network enclave
  - Directs transfers through ***single network ingress/egress point*** that can be ***monitored*** (e.g., with the Bro IDS) and ***restricted*** (e.g., with router ACLs)
  - Achieves throughput by ***reducing complexity***


Border Router
Enterprise Border Router/Firewall
WAN
10G
10GE
perfSONAR
Clean, High-bandwidth WAN path
Site / Campus access to Science DMZ resources
10GE
10GE
Site / Campus LAN
Science DMZ Switch/Router
10GE
perfSONAR
Per-service security policy control points
High performance Data Transfer Node with high-speed storage
🟩 High Latency WAN Path
🟥 Low Latency LAN Path

# Medical Science DMZ



## The medical science DMZ

RECEIVED 2 December 2015
REVISED 17 January 2016
ACCEPTED 11 February 2016

Sean Peisert, PhD[1,2], William Barnett, PhD[3], Eli Dart, BS[4], James Cuff, D.Phil[5], Robert L Grossman, PhD[6], Edward Balas, BS[7], Ari Berman, PhD[8], Anurag Shankar, PhD[9], Brian Tierney, MS[4]

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD
UNIVERSITY PRESS

- Applies Science DMZ framework to computing environments requiring compliance with HIPAA Security Rule

- Key architectures:
  - All traffic from outside compute/storage infrastructure passes through heavily monitored “head nodes.”
  - Storage/compute nodes are not connected directly to the Internet.
  - Traffic containing sensitive or controlled access data is encrypted.

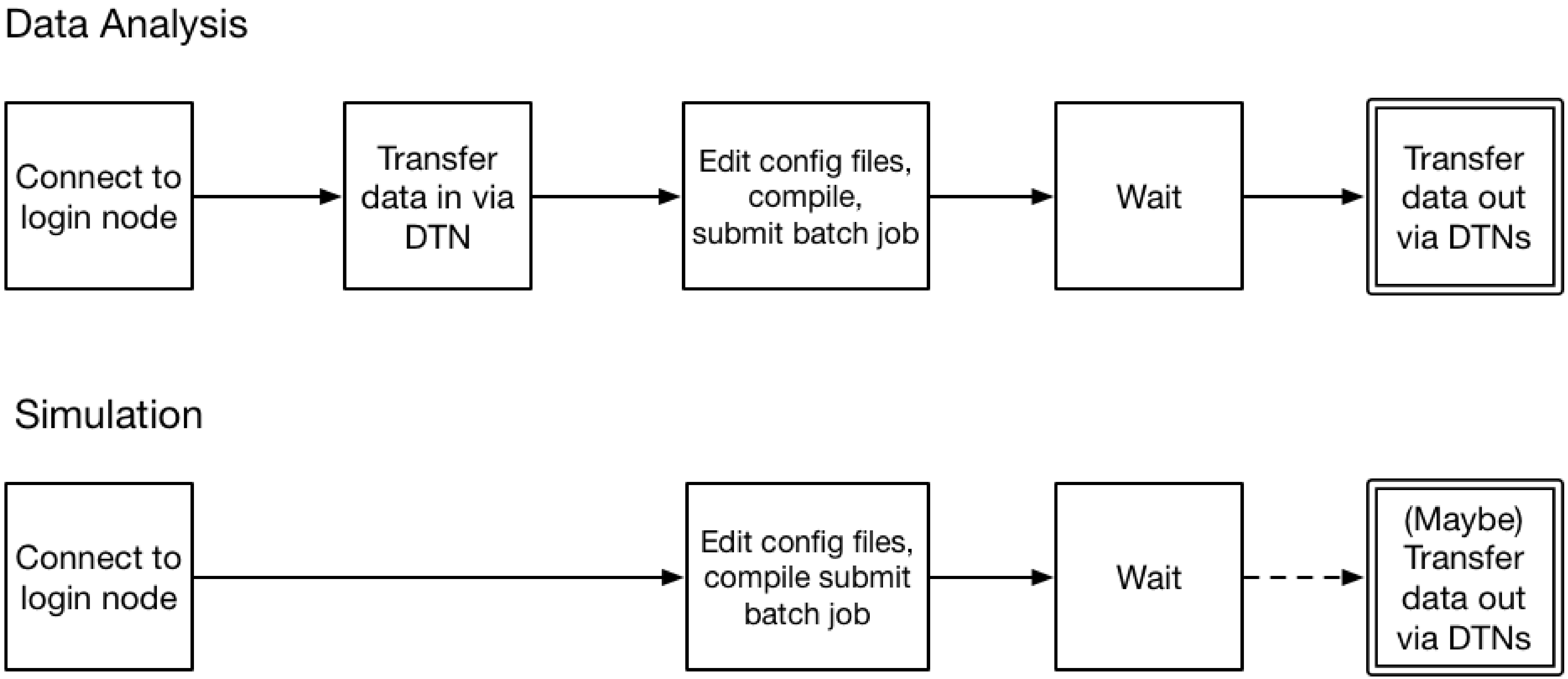
The Science DMZ helps compensate for HPC’s limitations — we need more such solutions.

We also need solutions that can leverage HPC distinctiveness as a strength.

# HPC Security Opportunities

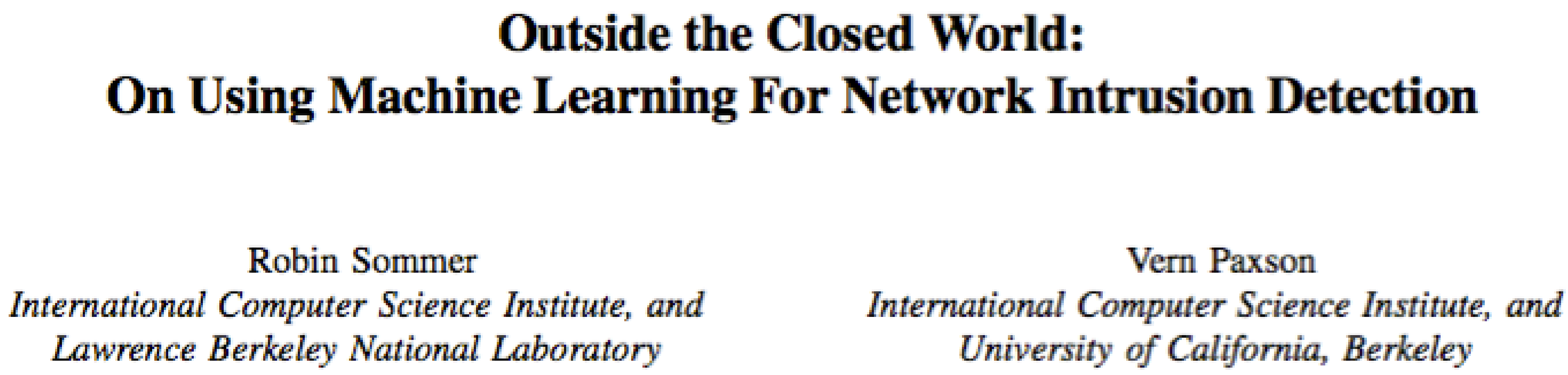
Key Point #1: HPC systems tend to be used for very distinctive purposes, notably mathematical computations

# HPC systems tend to be used for very distinctive purposes, notably mathematical computations


Data Analysis
Connect to login node
Transfer data in via DTN
Edit config files, compile, submit batch job
Wait
Transfer data out via DTNs
Simulation
Connect to login node
Edit config files, compile submit batch job
Wait
(Maybe) Transfer data out via DTNs

How do we leverage this insight for security?

# Intrusion Detection



**Outside the Closed World:**
**On Using Machine Learning For Network Intrusion Detection**

Robin Sommer
*International Computer Science Institute, and*
*Lawrence Berkeley National Laboratory*

Vern Paxson
*International Computer Science Institute, and*
*University of California, Berkeley*

"…machine learning is rarely employed in operational "real world" settings. … task of finding attacks is fundamentally different from … other applications,

"… Network traffic often exhibits much more diversity .. which leads to misconceptions about what anomaly detection … can realistically achieve…"

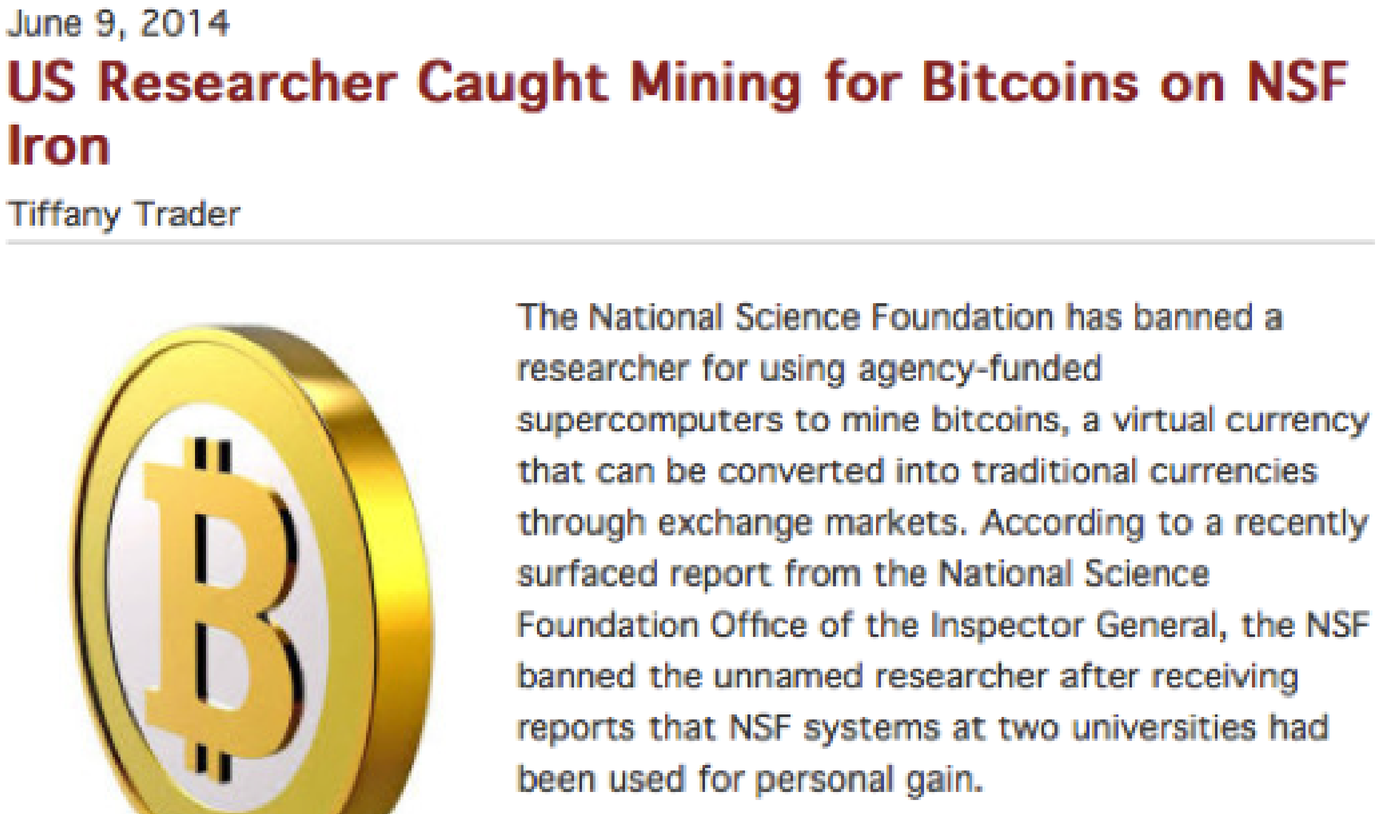
"… we argue for the importance of … insight into … an anomaly detection system from an operational point of view. It is crucial to acknowledge [the difficulty in making] progress … without any semantic understanding…"

R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," *Proc. 31st IEEE Symposium on Security & Privacy*, May 2010.

Key Point #2: What if there was less diversity in the activities we monitor, and greater semantic understanding?

# Fingerprinting Computation on HPC Systems

- What are people running on HPC systems?
  - Are they running what they usually run?
  - Are they running what they requested cycle allocations to run?
  - Are they running something illegal (e.g., classified?)



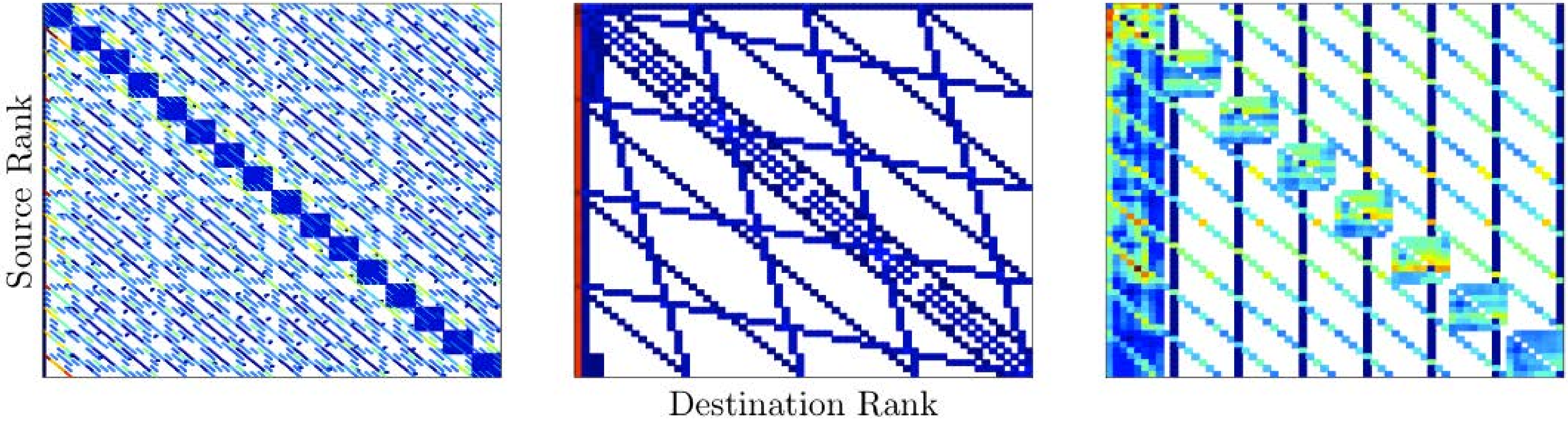
June 9, 2014

**US Researcher Caught Mining for Bitcoins on NSF Iron**

Tiffany Trader

The National Science Foundation has banned a researcher for using agency-funded supercomputers to mine bitcoins, a virtual currency that can be converted into traditional currencies through exchange markets. According to a recently surfaced report from the National Science Foundation Office of the Inspector General, the NSF banned the unnamed researcher after receiving reports that NSF systems at two universities had been used for personal gain.

# Adjacency Matrices of Communication Patterns



Adjacency matrices for individual runs of a performance benchmark, an atmospheric dynamics simulator, and a linear equation solver SUPERLU (64 nodes).

Number of bytes sent between ranks is linearly mapped from dark blue (lowest) to red (highest), with white indicating an absence of communication

# Fingerprinting Computation on HPC Systems

- Developed technique for fingerprinting communication on HPC systems
- Experiments & Results
  - Used 1681 logs for 29 scientific applications from NERSC HPC systems
  - Applied Bayesian-based machine learning technique for classification of scientific computations.
  - Applied graph-theoretic approach using “approximate” graph matching techniques (subgraph isomorphism & edit distance)
  - Hybrid learning / graph theory approach identifies test HPC codes with 95-99% accuracy.

Network-Theoretic Classification of Parallel Computation Patterns,” *International Journal of High Performance C*

Bishop, “Multiclass Classification of Distributed Memory Parallel Computations,” *Pattern Recognition Letters*, 3

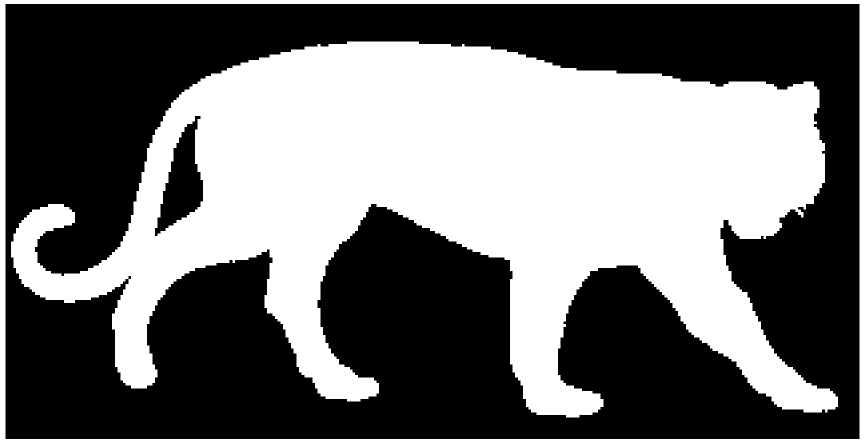
# HP Security Opportunities: Monitoring Data

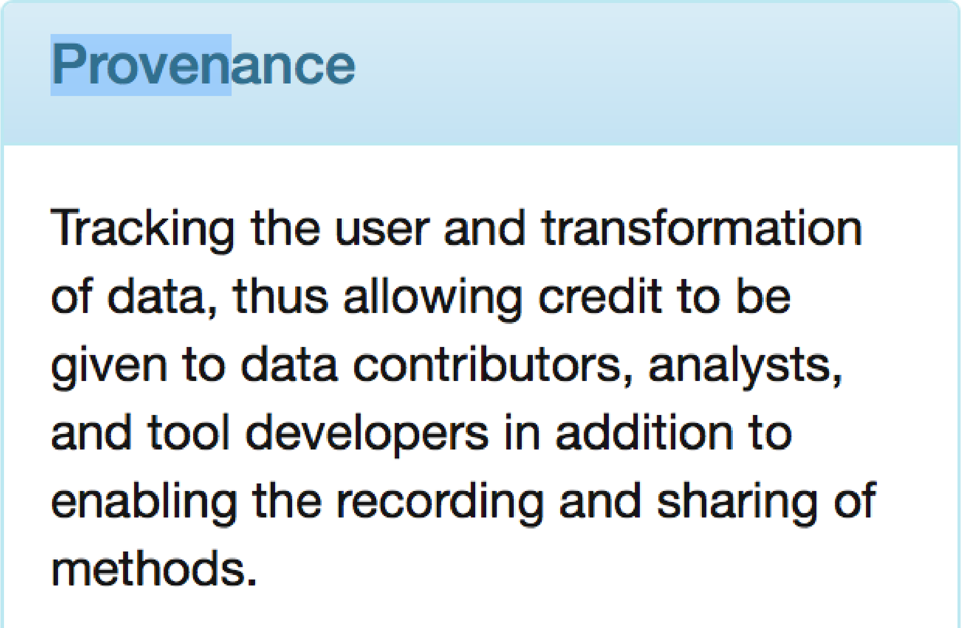
- Monitoring data is useful for *security monitoring for abnormal behavior*
  - Misuse of cycles
  - Identifying manipulated programs (malware, etc..)

- Monitoring data is also useful for *provenance / integrity monitoring*

- Ability to successfully perform analysis on monitoring data depends on *availability of useful monitoring data*

# Current focus on provenance in HPC might help provide better monitoring data





Provenance

Tracking the user and transformation of data, thus allowing credit to be given to data contributors, analysts, and tool developers in addition to enabling the recording and sharing of methods.

AURORA

mOS

INTEL® OMNI-PATH FABRIC 100 SERIES

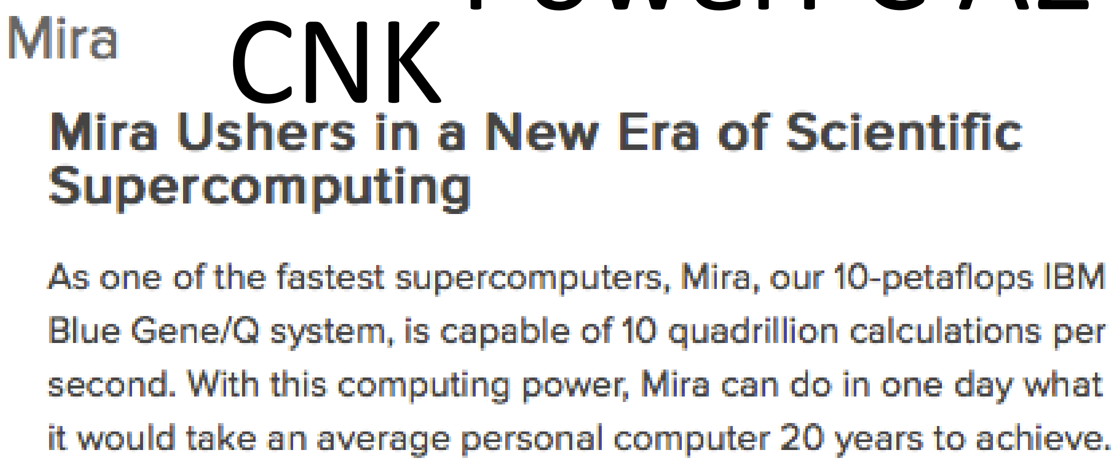
CNL

HPC systems that run exotic hardware and software stacks may also provide monitoring data

Lustre®

GPUs

PowerPC A2


Mira

CNK

**Mira Ushers in a New Era of Scientific Supercomputing**

As one of the fastest supercomputers, Mira, our 10-petaflops IBM Blue Gene/Q system, is capable of 10 quadrillion calculations per second. With this computing power, Mira can do in one day what it would take an average personal computer 20 years to achieve.

SUMMIT

NVLINK HIGH-SPEED INTERCONNECT
Designed for Accelerated Computing

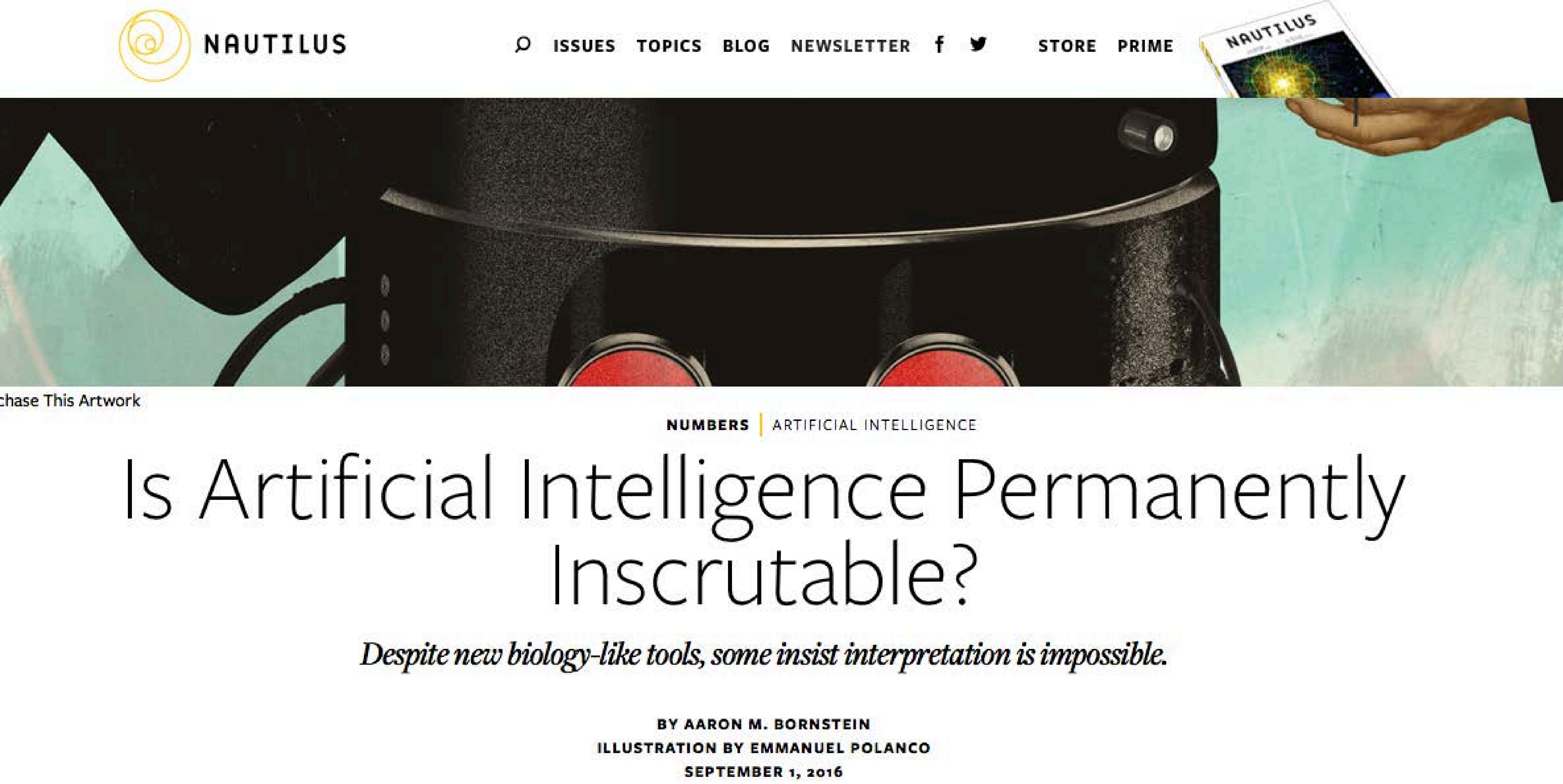
U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

# Future/Alternative Architectures May Continue This

- GPU-based computation
- Exascale proxies
- Neuromorphic chips
- Quantum computing

# Future/Alternative Architectures May Continue This



NUMBERS | ARTIFICIAL INTELLIGENCE

## Is Artificial Intelligence Permanently Inscrutable?

*Despite new biology-like tools, some insist interpretation is impossible.*

BY AARON M. BORNSTEIN
ILLUSTRATION BY EMMANUEL POLANCO
SEPTEMBER 1, 2016

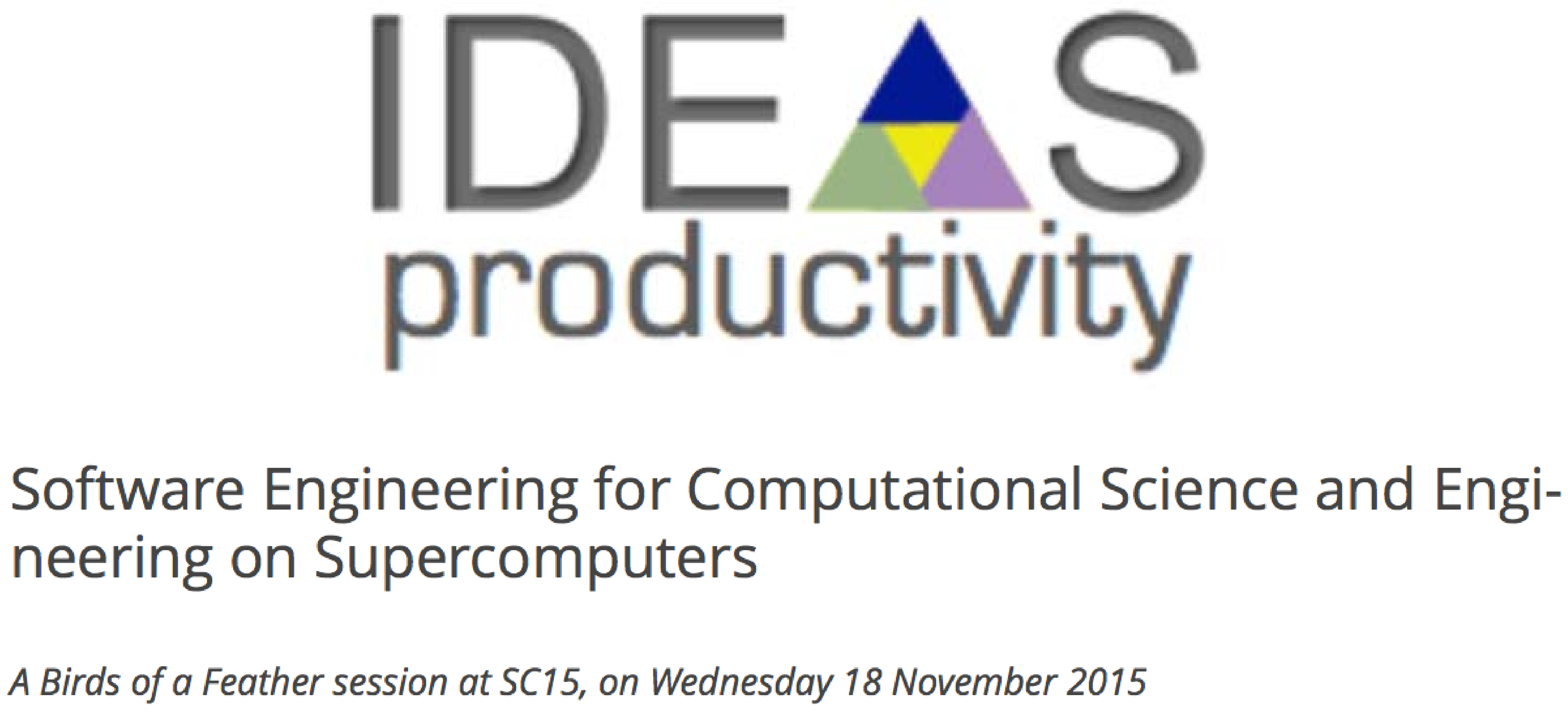
# Exotic hardware and software stacks in HPC systems

- Partially a liability:
  - These stacks may be tested less thoroughly as larger stacks

- Partially an opportunity:
  - Key point #3: *some custom stacks may be smaller, and more easily verified.*
  - Key Point #4: *custom stacks provide opportunities for instrumenting system hardware or software to capture additional audit/provenance data.*

# Looking to the future

# Software engineering is a key goal of the NSCI



Software Engineering for Computational Science and Engineering on Supercomputers

*A Birds of a Feather session at SC15, on Wednesday 18 November 2015*

- *Automated static/runtime analysis tools might be developed to check HPC code for insecure behaviors.*

# Looking forward: science is changing

- *Challenges*
  - Sensor data
  - Distributed / streaming data collection

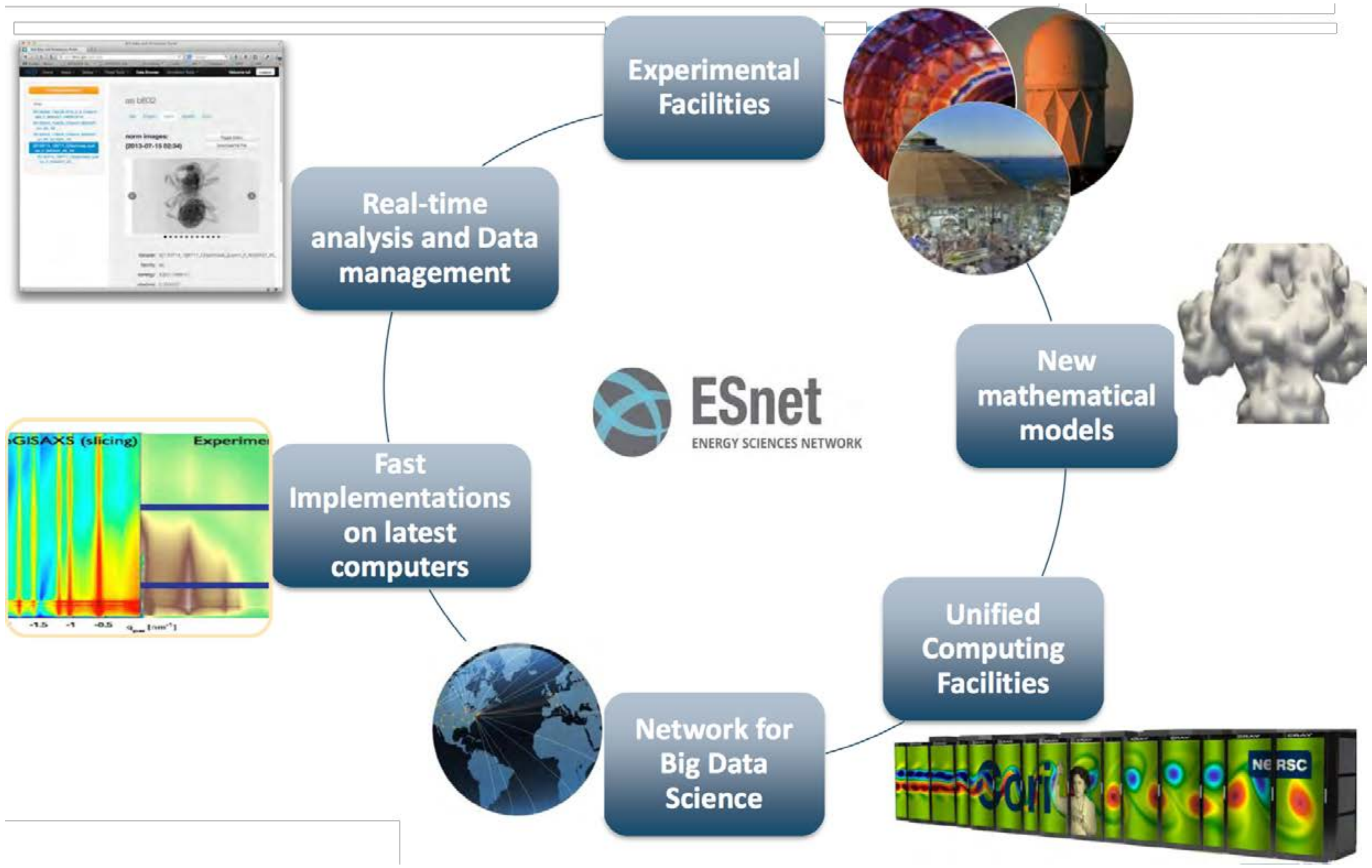
- *Science data is getting to us in new ways, and we have more data to protect.*

# Trend toward constrained modes of operation

- *Containerization* — all interaction takes place within the container

# Trend toward constrained modes of operation

- *Limited interfaces / “Automated Supercomputing”*
  - Science gateways — web portals to HPC
  - “Superfacility” model


Experimental Facilities
Real-time analysis and Data management
New mathematical models
ESnet
ENERGY SCIENCES NETWORK
GISAXS (slicing)
Experime
Fast Implementations on latest computers
Unified Computing Facilities
Network for Big Data Science
NERSC

# Future opportunities & challenges with “constrained” operation

- *Opportunities:*
  - Security tends to benefit from more constrained operation, which is the general trend.

- *Challenges:*
  - Still have vulnerabilities from arbitrary code (even when it’s submitted via web front-ends)

# Looking forward to “future” security technologies

- Differential privacy
- (Quasi)-homomorphic encryption (e.g,. CryptDB/Mylar)
- In combination with “limited” interfaces to raw data being used, may provide new means for interacting with data sets in *constrained* fashion.

# Open research questions

- *How can we better understand behaviors in HPC systems and the effect of those behaviors on security/scientific computing integrity?*

    - understand where monitoring data can be *collected* in HW & SW,
    - *how to analyze that data* to evaluate security-related behaviors,
    - *value of that data* for computer security and/or scientific computing integrity verification purposes, and
    - how to do this *efficiently* (minimal collection overhead, and near-real-time analysis)
    - understand ways in which *behaviors on HPC systems and system complexity can be reduced* to enable better analysis

# Summary

- HPC systems:
  - some things similar to ordinary IT computing
  - some significant differences — *challenges and opportunities*.

- Key security challenges:
  - Traditional security solutions often aren't effective given *priority of high-performance*.
  - Many HPC environments are *highly "open"* to enable broad scientific collaboration.

- Key security opportunities:
  - HPC systems tend to be used for very *distinctive purposes*, notably mathematical computations.
    - The *"regularity" of activity* within HPC systems can benefit the effectiveness of *machine learning analyses on security monitoring data* to detect *misuse of cycles* and *threats to computational integrity*.
  - *Custom HW/SW stacks* provide opportunities for enhanced *security monitoring*.
  - Trend toward ***containerized operation*** & ***limited interfaces*** in HPC is likely to help.

*Contact:*

*Dr. Sean Peisert*

*sppeisert@lbl.gov*

*http://crd.lbl.gov/Q/peisert/*

U.S. DEPARTMENT OF ENERGY

UNIVERSITY OF CALIFORNIA

BERKELEY LAB
Lawrence Berkeley National Laboratory