Thanks

- Pratul Agarwal, Ryan Doll, Terrance Figy, and entire OAK/Wichita State team!
- Yang Guo and NIST for co-support
NSF: What We Do

• **Discovery**: NSF supports U.S. researchers to generate new knowledge and discoveries that transform the understanding of the world, while also transforming modern society through technological innovations. Situated at the intersection of all S&E disciplines, NSF is also uniquely positioned to identify and guide investments toward emerging frontier areas for scientific research.

• **Research Infrastructure**: NSF funds supercomputers, ground-based telescopes, U.S. research stations in the Arctic and Antarctic, the world’s largest and highest-powered magnet lab, long-term ecological sites, engineering centers and other infrastructure and state-of-the-art tools to sustain the nation’s scientific enterprise.

• **Learning**: NSF programs support STEM education and training that attract individuals from every sector and group in society, ensuring a pipeline of people and ideas ready to solve the pressing global challenges in STEM.
NSF Directorates and Science

- Engineering
- Computer & Information Science & Engineering
- Geosciences (including Polar Programs)
- Education & Human Resources
- Integrative Activities
- Mathematical & Physical Sciences
- Biological Sciences
- Social, Behavioral & Economic Sciences
- International Science & Engineering
OAC: Transforming scientific discovery through cyberinfrastructure
NSF’s CISE/OAC and Scientific Cyberinfrastructure

• **Cyberinfrastructure (CI):** Compute, data, software, networking and people to facilitate scientific discovery and innovation.

• **Office of Advanced Cyberinfrastructure:** Supports and coordinates the development, acquisition and provisioning of state-of-the-art cyberinfrastructure resources, tools and services essential to the advancement and transformation of science and engineering.
<table>
<thead>
<tr>
<th>OAC investment areas</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Computing</td>
<td>Production and operational level advanced computing and data capabilities and services</td>
</tr>
<tr>
<td>Networking &amp; Cybersecurity</td>
<td>Advanced capabilities that preserve security and privacy</td>
</tr>
<tr>
<td>Learning &amp; Workforce Development</td>
<td>Foster a national research workforce for creating, utilizing, and supporting advanced CI</td>
</tr>
<tr>
<td>Software &amp; Data CI</td>
<td>Develop a cohesive, federated, national-scale approach to research data infrastructure</td>
</tr>
<tr>
<td>Strategic Investments</td>
<td>Special opportunities, cross-cutting and national initiatives, CI for open science and public access</td>
</tr>
</tbody>
</table>
### OAC investment areas

<table>
<thead>
<tr>
<th>Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advanced Computing</strong></td>
<td>Production and operational level advanced computing and data capabilities and services</td>
</tr>
<tr>
<td><strong>Networking &amp; Cybersecurity</strong></td>
<td>Advanced capabilities that preserve security and privacy</td>
</tr>
<tr>
<td><strong>Learning &amp; Workforce</strong></td>
<td>Foster a national research workforce for creating, utilizing, and supporting advanced CI</td>
</tr>
<tr>
<td><strong>Software &amp; Data CI</strong></td>
<td>Special opportunities, cross-cutting and national initiatives, CI for open science and public access</td>
</tr>
<tr>
<td><strong>Strategic Investments</strong></td>
<td>Cannot realize science goals unless cyberinfrastructure is secure, robust, and trustworthy.</td>
</tr>
</tbody>
</table>
OAC CI Cybersecurity Vision

NSF’s Blueprint for a National CI Ecosystem for the 21st Century: “Agile, integrated, robust, trustworthy, and sustainable CI ecosystem that drives new thinking and transformative discoveries in all areas of S&E research and education”
USC/ISI SPHERE: Research Infrastructure for Cybersecurity Experimentation

- **Heterogeneity**: cover 90% of research need: CPU, GPU, TEE, PLC, FPGA, IoT
- **Reproducibility**: built-in facilities; work with artifact evaluators
- **Realism**: at-scale, experimental composability, interfaces to public Internet, real traffic
- **Usability**: multiple user “portals” catering to different levels of need and experimental sophistication
- **Participation**: cater to education and research
CI resources and services for the research community

Democratized access to advanced computing

- ACCESS: https://access-ci.org/
- LCCF: https://lccf.tacc.utexas.edu/
- PaTh: https://path-cc.io/
- SGX3: https://sciencegateways.org/
- MSCC: https://www.ms-cc.org/
- RCD Nexus: https://rcd-nexus.org/
- Trusted CI: https://www.trustedci.org/
- Research SOC: https://omnisoc.iu.edu/services/researchsoc/
- CI Compass: https://ci-compass.org/

CI services for NSF major and mid-scale RI

- TRUSTED CI: https://trustedci.org/
- ResearchSOC: https://researchsoc.org/

Community and workforce development

- CloudBank: Commercial cloud
- SGX3: Science Gateways expertise
- RCD Nexus: CI Workforce Development
- Minority Serving CI Consortium (MS-CC)

Facility data lifecycle

- Cybersecurity framework
- Security Operations
- Regulated Research

Portals:
OAC Updates
Changing user, technology, vendor and national landscape requires us to think deeply about our collective strategy for the future

- New user communities requiring computing and data infrastructure
- New technologies, hardware specialization, slowing of Moore’s law, IAAS and SAAS
- Rise of massive data and AI
- New business models and entrants into the ecosystem
- New and pending legislation and initiatives
National AI Research Resource (NAIRR)

Objective and Goals

Objective: To strengthen and democratize the U.S. AI Innovation ecosystem in a way that protects privacy, civil rights, and civil liberties

Goals:

- Spur innovation
- Increase the diversity of talent in AI
- Improve U.S. capacity for AI R&D
- Advance trustworthy AI

The NAIRR should comprise a federated set of computational, data, testbed, and software resources from a variety of providers, along with technical support and training,
Vision for the National AI Research Resource

A widely-accessible, national research infrastructure that will advance the U.S. AI R&D environment, discovery, and innovation by empowering a diverse set of users through access to:

- Secure, high-performance, privacy-preserving computing
- High-quality datasets
- Catalogs of testbeds and educational materials
- Training tools and user support mechanisms
**NAIRR Pilot Organization**

**User Journey**
- **US-based Researchers, Educators & Students**
  - Enable open AI research and access to diverse AI resources via a central portal and coordinated allocations

**Portal**
- https://nairrpilot.org

**Pilot Resources and Opportunities**
- The NAIRR Pilot provides infrastructure and resources; it does not fund end-user research.

**Operations**
- **NAIRR Open**
  - Enable open AI research and access to diverse AI resources via a central portal and coordinated allocations

- **NAIRR Secure**
  - Enable AI research needing privacy and security-preserving resources. Assemble exemplar privacy preserving resources.

- **NAIRR Software**
  - Facilitate use of AI software, platforms, tools and services across platforms

- **NAIRR Classroom**
  - Reach new communities through education, training, user support and outreach

**Governance**

**Community Design Process**

**NAIRR Pilot**
OAC People Initiatives

• Workforce:
  • CI (and CI cybersecurity) depends on people with specialized skills
  • Recognized need to build and grow pipeline of CI professionals

• Research:
  • Expand access to advanced CI
NSF-wide Strategy for CI Professionals

• Promote professional development, career paths, incentivize coordination; address sustainability
  • Nurturing Diverse, Skilled, Capable, and Productive Communities of Cyberinfrastructure Professionals (DCL; NSF 22-052)
    • CI Professional Mentoring and/or Professional Development Plan requirement in solicitations funding CI professionals
  • Better Scientific Software Fellows (https://bssw.io, partnership with DOE)

• Establish, foster, and nurture a community
  • CI CoE Pilot: Minority Serving Cyberinfrastructure Consortium (MSCC)
  • Research Coordination Networks: Fostering and Nurturing a Diverse Community of CI Professionals (RCN:CIP;-NSF 22-558)
  • Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining; NSF 22-574; due May 16, 2022)

• Develop academic structures/career paths
  • CI CoE Pilot: Research Computing and Data Resource and Career Center (http://rcd-nexus.org)
RCD Nexus: Supporting the Professionals who advance Computational and Data-intensive Research

Tools, Practices, and Professional Development Resources

- RCD Capabilities Model v2.0 and Data Exploration Portal
- RCD Professional Staffing survey - Who, How many, etc.
- Advance adoption of HR Framework for RCD Job Families/classifications
- Gather and share Leading Practices for Staff Recruitment, Retention, and Professional Development, as well as Student Workforce Development
- Document Career Arcs to inform hiring, training, & career options

Gathering the Community of Communities

- Foster connections among the communities supporting RCD professionals
- Collaborate to develop a shared voice to advocate for this new profession
- Work together to increase diversity, equity, and inclusion

Supported by NSF OAC 210003  http://rcd-nexus.org
EPSCoR
Established Program to Stimulate Competitive Research (EPSCoR)

Mission
Enhance research competitiveness of targeted jurisdictions by strengthening STEM capacity and capability

Goals
- Catalyze research capability across and among jurisdictions
- Establish STEM professional development pathways
- Broaden participation of diverse groups and institutions in STEM
- Effect engagement in STEM at national and global levels
- Impact jurisdictional economic development
Engaging EPSCoR jurisdictions is critical

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>18%</td>
<td>total US population</td>
</tr>
<tr>
<td>24%</td>
<td>nation’s accredited universities</td>
</tr>
<tr>
<td>23%</td>
<td>nation’s Emerging Research Institutions</td>
</tr>
<tr>
<td>30%</td>
<td>nation’s MSIs</td>
</tr>
<tr>
<td>50%</td>
<td>nation’s HBCUs</td>
</tr>
<tr>
<td>29%</td>
<td>nation’s HSIs</td>
</tr>
<tr>
<td>69%</td>
<td>nation’s TCUs</td>
</tr>
<tr>
<td>10%</td>
<td>nation’s AANAPISIs</td>
</tr>
<tr>
<td>19%</td>
<td>nation’s African Americans</td>
</tr>
<tr>
<td>18%</td>
<td>nation’s Hispanics</td>
</tr>
<tr>
<td>39%</td>
<td>nation’s American Indians</td>
</tr>
<tr>
<td>44%</td>
<td>nation’s Pacific Islanders</td>
</tr>
</tbody>
</table>
**EPSCoR jurisdictions vary** in metrics including proposals submitted, institutions, awards, and obligations (FY19-22)

<table>
<thead>
<tr>
<th>States by Metrics</th>
<th>Quartile 1</th>
<th>Quartile 2</th>
<th>Quartile 3</th>
<th>Quartile 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alabama</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guam</td>
<td>$15.96M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idaho</td>
<td></td>
<td>$122.75M</td>
<td></td>
<td>$177.27M</td>
</tr>
<tr>
<td>Kansas</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td></td>
<td>$116.45M</td>
<td></td>
<td>$168.54M</td>
</tr>
<tr>
<td>Mississippi</td>
<td>$152.38M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td></td>
<td>$73.51M</td>
<td></td>
<td>$101.93M</td>
</tr>
<tr>
<td>Nebraska</td>
<td></td>
<td></td>
<td>$152.42M</td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td></td>
<td></td>
<td>$156.06M</td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>$119.22M</td>
<td></td>
<td></td>
<td>$160.47M</td>
</tr>
<tr>
<td>Nevada</td>
<td></td>
<td>$80.78M</td>
<td></td>
<td>$158.90M</td>
</tr>
<tr>
<td>Oklahoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puerto Rico</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td></td>
<td>$84.03M</td>
<td>$212.68M</td>
<td>$257.96M</td>
</tr>
<tr>
<td>South Carolina</td>
<td>$29.15M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>$49.27M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virgin Islands</td>
<td>$67.21M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td></td>
<td></td>
<td></td>
<td>$92.68M</td>
</tr>
</tbody>
</table>

*Data from NSF by the Numbers, accessed 8/31/23.*
Key NSF EPSCoR Highlights from CHIPS & Science Act

(SEC. 10325: EXPANDING GEOGRAPHIC AND INSTITUTIONAL DIVERSITY IN RESEARCH)

- Authorization of a gradual increase in funding for institutions in EPSCoR jurisdictions.

<table>
<thead>
<tr>
<th>FY23</th>
<th>FY24</th>
<th>FY25</th>
<th>FY26</th>
<th>FY27</th>
<th>FY28</th>
<th>FY29</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.5%</td>
<td>16%</td>
<td>16.5%</td>
<td>17%</td>
<td>18%</td>
<td>19%</td>
<td>20%</td>
</tr>
</tbody>
</table>

- Authorization of a gradual increase in funding of scholarships, graduate fellowships and traineeships, and postdoctoral awards to support EPSCoR institutions.

<table>
<thead>
<tr>
<th>FY23</th>
<th>FY24</th>
<th>FY25</th>
<th>FY26</th>
<th>FY27</th>
<th>FY28</th>
<th>FY29</th>
</tr>
</thead>
<tbody>
<tr>
<td>16%</td>
<td>18%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>
EPSCoR Investment Strategies

Research Infrastructure Improvement (RII) (78-84% of EPSCoR budget)
+ Support physical, human, and cyber infrastructure within academic institutions across each jurisdiction

Co-Funding w/ NSF Directorates & Offices (16-22% of budget)
+ Meritorious proposals reviewed in other NSF programs that also satisfy EPSCoR programmatic criteria

Outreach and Workshops (0.5-1% of budget)
+ Interaction among EPSCoR Community and NSF to build mutual awareness and develop areas of potential strength
Keeping state and momentum

- NIST NCCoE, Gaithersburg, MD
- ~100 attendees:
  - Operators
  - Researchers
  - Government
  - Industry

NIST Interagency Report
NIST IR 8476

3rd High-Performance Computing Security Workshop
Joint NIST-NSF Workshop Report
Last year: RIG Cybersecurity Thematics

- Explicit acknowledgement of individual facility uniqueness and requirements:
  - “The foundation for developing and maintaining a project’s cybersecurity program lies in the research mission and goals of the facility itself”

- Incentivize cybersecurity rather than mandate / regulate / audit
  - Carrots vs. sticks: provide supporting resources that benefit cyberinfrastructure, facility, and scientific discovery mission

- Living document:
  - As cybersecurity techniques, tools, and threats evolve, so too do the guidelines
<table>
<thead>
<tr>
<th>Researcher Dilemma</th>
<th>Results</th>
<th>Government and Sponsor Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>You are focused on research, not cybersecurity (as it should be)</td>
<td>Compromise of research data and instruments, impacting efficiency, trustworthiness, reproducibility, and funding</td>
<td>Increasingly complex cybersecurity requirements in grants, contracts, and data use agreements</td>
</tr>
<tr>
<td>It is challenging to translate cybersecurity into implementables</td>
<td></td>
<td>New regulations</td>
</tr>
</tbody>
</table>
Framing and Urgency

• NSF HPC increasingly considered critical infrastructure / major infrastructure
• National competitiveness and reputation can be put at risk (even w/o loss of data)
• Incidents draw attention of:
  • The press
  • Federal agencies
  • Executive offices
• Results in increased pressure on NSF to become prescriptive, creating a vicious cycle
RIG Directions

• Updates coming in 2025; under purview of NSF’s Research Infrastructure Office (RIO)
• Theme: cybersecurity is risk management, requires leadership engagement, building resilience
• Possible additions / reporting requirements:
  • Cyber risk register
  • Cybersecurity budgets
  • Information assurance management plan
How is NSF HPC Security Unique

- Novel architectures
  - E.g., neuromorphic, quantum, experimental
- Specialized software, workloads, and data
  - Users bring own code
- Unique users
  - Highly collaborative / distributed
  - International
- Performance focus
  - Availability secondary
- Science mission
  - Trust in science
- Different adversaries
  - Open data
A (few) challenges from 2023

• Operators:
  • New access methods (beyond ssh/bastion)
  • Host homogeneity (feature and bug)
  • MFA w/ remote / non-institutional users?
  • Dynamic workloads
  • User-installed / compiled code
  • Supply chain security
  • Virtualization / containers
  • Compliance (NIST RMF)
  • “Protect the science”
OAC Supported HPC Security Research
Whither Cybersecurity: We do open and unclassified science!”

- How can cybersecurity benefit the cyberinfrastructure?
- Imagine a world where...
  - Data has strong integrity protection, to prevent accidental or malicious modification
  - Research artifacts contain provenance meta-data
  - Collaboration between scientists and infrastructure is seamless and natural
  - Computation and sharing of sensitive data is possible without compromising privacy
  - Infrastructure is highly available and not vulnerable to mis-use
  - Third-parties can replicate and reproduce research findings
  - The public trusts science
NSF 23-517: Cybersecurity Innovation for Cyberinfrastructure (CICI)

- The objective of the CICI program is to develop, deploy and integrate solutions that benefit the broader scientific community by securing science data, workflows, and infrastructure.

<table>
<thead>
<tr>
<th>Applied research to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usable and Collaborative Security for Science <em>(UCSS)</em></td>
</tr>
<tr>
<td>Reference Scientific Security Datasets <em>(RSSD)</em></td>
</tr>
<tr>
<td>Transition to Cyberinfrastructure Resilience <em>(TCR)</em></td>
</tr>
</tbody>
</table>
Traffic Light Protocol

- System of markings that designates the extent to which recipients may share potentially sensitive information
- Used (to great effect) at NSF TrustedCI 2023 Cybersecurity summit

Here are the 4 LABELS

- **TLP:RED**
  - LIMITED TO RECIPIENT ONLY - You can act on a TLP:RED cybersecurity document if you receive one, but you must not convey it to anyone else.

- **TLP:AMBER**
  - LIMITED DISCLOSURE - This information can only be shared on a need-to-know basis among those within your organization and its customers.
  - The source may restrict sharing to the organization by setting TLP:AMBER+STRICT.

- **TLP:GREEN**
  - LIMITED DISCLOSURE TO COMMUNITY - You may share this information within your community. The TLP leaves it up to you to be reasonable about which people constitute your community.

- **TLP:CLEAR**
  - DISCLOSURE IS NOT LIMITED - Recipients can share this information with everyone.

---

[TLP:RED] How we failed to handle a triple-combo attack against the R&E HPC community worldwide...in the middle of a pandemic (In-Person Only)

(Romain Wartel)

Auditorium-50
CICI:TSR: Improving the Robustness of Cyberinfrastructure via Scalable Vulnerability Discovery and Mitigation on “Big Binaries”

Project Summary
Software upholds both the modern society and critical scientific cyberinfrastructure. Software powering scientific cyberinfrastructure often appears in the form of binaries and lacks maintenance and security practices. In past decades, emerging binary analysis techniques have changed how we analyze binary programs. We need significant technology transition effort to identify, implement, and evaluate these techniques using a robust binary analysis framework on a comprehensive corpus of binary programs, covering critical and legacy scientific software. The outcome will be a Cyber Reasoning System (CRS) for automatically finding and mitigating vulnerabilities in legacy binaries, and an open and comprehensive corpus of legacy binaries commonly seen in scientific software and cyberinfrastructure.

Planned Deliverables
We plan to produce a Cyber Reasoning System (CRS) that is capable of finding and mitigating vulnerabilities in legacy binary programs. We will also produce a comprehensive corpus of legacy binaries that are commonly seen in scientific software and cyberinfrastructure. We will release both the CRS and our redistributable binary corpus to the public with a permissive license.

Scientific and Broader Impacts
This project will provide novel and robust means to discover vulnerabilities in complex and real-world legacy binaries that are prevalent in scientific settings and cyberinfrastructure. Additionally, the binary corpus will be a large-scale and objective data set for evaluating future binary analysis techniques. It will help improve the security of scientific cyberinfrastructure, the cyber world, our society, and the nation as a whole.

This project will also benefit computer security education. We plan to create a course at ASU, “Automated Binary Code Analysis,” to teach our students how to create and tweak binary analysis techniques and apply them, in practice and at scale, on real-world software.

Research Challenges
• Building comprehensive corpora of legacy scientific binaries
• Building a flexible and scalable CRS
• Integrating state-of-the-art vulnerability discovery and mitigation techniques

Contact Information
• Ruoyu “Fish” Wang, fishw@asu.edu
• Yan Shoshitaishvili, yans@asu.edu

Ex 1: Bespoke Code
• Written in legacy languages
• Maybe unmaintained
• Maybe source not even available
• How to ensure security?
• Binary vulnerability analysis
## Motivation and Overview
- Detecting configuration-related software vulnerabilities in high-performance computing (HPC) systems is difficult due to the highly configurable environments.
- State-of-the-art bug detectors cannot solve this problem because they do not take into account the specialized HPC contexts of interdependent software components.
- Connecting analysis to contexts is extremely hard in generic settings, as the combination of hardware and software resources varies greatly.
- This project develops **deployment-specific vulnerability detection** that leverages unique HPC characteristics to facilitate the discovery of configuration errors.

## Technical Approach
- Study the deployment contexts in real-world high-performance computing systems and develop both offline and online tools to automatically collect such contextual information.
- Apply extracted contexts to detecting misconfiguration and configuration-triggered code vulnerabilities at both deployment time and incrementally at runtime.
- Test the novel tools in real-world testbeds and high-performance computing environments to evaluate their accuracy, efficiency and effectiveness.

## Intellectual Merit
- **Novel Insights**: HPC deployments are highly configurable and software vulnerabilities originate from specific deployment contexts.
- **New and HPC-Specific Approach**: This research takes full advantage of the *de facto* workflow of high-performance computing systems, so as to make it possible to enable vulnerability detection in a deployment context-aware manner.
- **Customized Techniques**: The understanding of high-performance computing contexts allows us to customize state-of-the-art analysis techniques and make them more efficient in this new domain.
- **Realistic Testbed**: This project builds a novel, comprehensive and realistic HPC security testbed that can facilitate the design, implementation and evaluation of these new techniques.

## Broader Impact
- Sharing the HPC security testbed with faculty, graduate researchers and undergraduate students who otherwise do not have access to such an interdisciplinary platform.
- Providing a comprehensive understanding of the software security problems in real-world scientific computing systems.
- Making a significant impact on the robustness of the national bottom line, as scientific computing are increasingly applied to critical areas in our society such as COVID-related research.
- Disseminating our code, data and publication to the public.
- A variety of educational activities including mentoring 10-12th graders in the GREAT-Advanced Robotics Camp at University of Utah, developing new graduate and undergraduate courses, etc.
CICI: UCSS: ARMOR: Secure Querying of Massive Scientific Datasets
PI: Hoda Maleki. co-PIs: Gagan Agrawal, Benjamin Fuller

Ex 3: Computing on Private Data

Problem and Motivation
- Scientific Shared Data is:
  - Massive, infrequently and sparsely accessed
  - A driver of new discovery
  - Sensitive and Private, represents a strategic advantage
- Natural approaches to protect scientific data in a shared cloud:
  - Policy, rely on cloud providers
  - Trusted hardware execution environment
  - General purpose cryptographic primitives with high overheads
  - Specialized searchable encryption techniques - THIS PROJECT

Key Outcomes
- New Encryption and Query Processing Techniques
  - Address challenges of equality-based selection, multidimensional range selection, and joining on value similarity or ranges queries.
- Scale and Efficiency Oriented Designs
  - Novel representation supporting joins and encryption
  - Integrate required parallelism with encryption.
- Demonstration of Low Overheads and Leakage in Science Contexts.
- New curriculum development and involvement of students from underrepresented groups.

Approach
- Combine multimap with search algorithm of Cash et al. and range covers.
- Augment Kamara and Moataz's construction with Boolean multimaps, property-preserving encryption, and m-out-of-n locality sensitive hash.
- Evaluate the security of our approach by current leakage attacks
  - Consider snapshot and persistent adversaries.
- Prove the security and privacy properties of our approach
- Prototype solutions, simulate data sizes and queries, and evaluate the overhead of storage, network bandwidth, and request/response times.
Prioritizing patches, understanding dependencies in HPC

Ex. 4: CICI: TCR: Prompt, Reliable, and Safe Security Update for Cyberinfrastructure

**Patch Presence Management**
- **Goal**: Identify available but unadopted patches in a prompt fashion for cyberinfrastructure.
- **Proposed Research**: Adapting open-source frameworks to provide affordable patch management and exploring patch presence detection and automated exploit generation to detect unadopted patches and assess their urgencies.

**Patch Safety Assessment**
- **Goal**: Understand the safety of an available patch when deployed to the target cyberinfrastructure.
- **Proposed Research**: Developing a system-wide dependency analysis to understand the components impacted by a target patch and assembling a low-cost testbed on the fly to assess the safety of the patch to the entire system.

**Patch Reliability Testing**
- **Goal**: Assess whether an available patch can reliably fix the target vulnerability without hurting the normal functionality.
- **Proposed Research**: Introducing directed fuzz testing and regression fuzz testing as methods to evaluate patch quality for cyberinfrastructure and tailoring differential program analysis to analyze the testing outcomes and measure the reliability of available patches.

**Technical Innovations**: Prompt, reliable, and safe security updates for cyberinfrastructure under challenging conditions, including limited monetary resources, insufficient admin expertise, and highly diverse environments.

**Broader Impacts**: Deployment to protect Utah CHPC (a 5,600-user platform) and possibly other cyberinfrastructure platforms (e.g., PNNL, ORNL, OARC), undergrad- and grad-level courses, underrepresented students’ education, and K-12 outreach.
Ex 5. HPC container security and minimization

CICI: UCSS: Secure Containers in High-Performance Computing Infrastructure

Problem:
• Goal: Designing secure containers for high-performance computing (HPC) infrastructures.
• Existing solutions are insecure:
  • Container images are insecure. E.g., a recent study on neuroscience container images shows that there are 460 vulnerabilities per image.
  • The weak isolation between containers and hosts can lead to vulnerabilities. We have observed 11 such vulnerabilities since 2017.

Research Overview:
• The proposed work in a simplified HPC infrastructure.

Thrust 1: Efficient vulnerability detection for container images
• Goal: Designing an efficient image vulnerability scanner to detect the images uploaded to the HPC infrastructure.
• Task 1-1 converts different types of code to embedded control flow graph (ECG).
• Task 1-2 converts ECG to code embedding with graph neural network and triplet-loss network.
• Task 1-3 proposes an efficient locality-sensitive hashing-based online search method.

Thrust 2: Secure, Lightweight and High-Performance Container Runtime
• Goal: Designing a container runtime tailored for the HPC infrastructure, which is both secure and high-performance.
• Task 2-1 uses a lightweight virtual machine hypervisor as the container runtime with various optimizations.
• Task 2-2 customizes the runtime based on HPC requirements on hypervisor feature, file system, network, and GPU.
• Task 2-3 designs a dynamical image debloating method that can remove unnecessary files, software, and packages.
Community Feedback

• Here to listen / learn from the broader HPC community
• We welcome your feedback and input!
  • What is OAC doing that’s working well?
  • What can we do better?
  • What should we be doing?
Thank You

“Make no little plans; They have no magic to stir men's blood …”
Daniel H. Burnham, Architect and City Planner Extraordinaire, 1907.

“If you want to travel fast, travel alone; if you want to travel far, travel together”
African Proverb.

Robert Beverly
Office of Advanced Cyberinfrastructure
rbeverly@nsf.gov

To subscribe to the OAC Announce Mailing List
Send an email to: OAC-ANNOUNCE-subscribe-request@listserv.nsf.gov