Practical Cybersecurity for Research Environments: Research Data, Infrastructure and Engagement Speed Learning Session

2019 NSF CC* and CICI PI Workshop

Speed Learning Facilitators: Von Welch, Indiana University, CACR; Tracy Futhey, Duke University

Guiding Questions for Group Discussions

What are some differences in needed functionality between enterprise IT environments and research computing environments?

What are some differences in risk between enterprise IT environments and research computing environments?

What are differences in risk between different types of research and who decides what level of risk can be accepted for the particular research project? For the university?

What are the biggest areas of friction between research and cybersecurity?
… research has urgency to 'get stuff done' and enterprise services and add delay and latency

Session 1 questions/discussion

Singularity containers… how to vet whether the image is safe to run?

How to deal with researcher storing data everyplace and insecurely?
... somebody's in charge of enterprise, everybody's in charge of research
… SDSU research data falls under enterprise classifications for data, so somewhat more solved than at some other universities
… GaTech cataloging data into an archival service (building now)

How to categorize 'harmless' vs. malicious data to build algorithms?

Where are people collecting data for network logs and willing to share..?
...UN-reno doing this and sharing
… science dmz traffic logging (even if deidentified) creates issues for looking at the data for practical cybersecurity

Who decides to accept the risk?
… most campuses, no explicit determination to accept risk researchers mostly oblivious of risk;
...SDSU - risk can be private/sensitive data, lack of backups, or intellectual property

Session 3….
.. pressures on openness of HPC systems vs. security requirements

...integrated security across projects with different requirements

… community account on their clusters on their accounts while still security for higher requiremetns (ITAR), etc

.. providing assurances at project level

Who decides on your campus about these risk tradeoffs
… CISO in supercomputing center _ CISO at Campus… communal effort and back and forth
… one campus was centralized (Nebraska), now at a new one that is famously decentralized (Wisconsin) and there good decisions are made within the levels, and with awareness of what is close to them. But how those connect is lacking

Risk management of enterprise vs.research
… generally more mature for administrative rather than research
… different use cases and different risk profiles and need to be managed differently

Distinction for researchers where data is the research vs. computing systems
… prioritization of critical assets is most important effort
… cost benefit is main consideration -- increase security without reducing access for researchers

Report Out: 3 Things

1. Difference between Enterprise and Research computing
    - SciDMZs, clusters, instruments
    - Many questions about securing various technologies
2. How risk questions are answered (Risk governance) for RC much less mature than Enterprise
    - Applying enterprise risk management to RC often bad
    - Stakeholders for research decentralized: PI, OSP/ORA, IRB, IT, etc.
3. Very different risks for different types of research
    - PHI vs HEP
    - Tensions e.g. community accounts versus 800-171
    - Increasing pressure against open research environments

Photos of Von’s Notes

Questions

1. How do deal with [non-]enterprise administration of Research HPC.?
2. Small institution - best practices for Res Cluster? How is it different?
3. Mapping DUA → CI?
4. Compliance issues. Who decides? How to manage?

Less cybersec, more Res Admin

3. IRB does talk to IT
   " " " Res Admin
   Decentralized!
   Connecting threads.
   Understanding Contracts
   Off Sponsored Programs
   Users doing stuff - we don't know it.

"Data is concen
not infrastructure"

↑ Bigger risk

must protect to protect data

(4) classifying data less rigid for research

Data lifecycle disposal?

Reproducibility

(2) Sci DMZs

Set of concerns
- Research + Normal staff

Port lock down by default
lots of exception for research

Templates for VMs
in collab w/ central IT

Faculty productivity loss is a risk

Balance?

Agility? > 1 hour? > 5 min. to unlock

Human in loop!

Do automated user mgmt