Emerging Research Platforms: Securing Sensitive Data from Lab to Likes

Workshop Report from
Camille Crittenden & Jeff Weekley

Workshop: Emerging Practices in Computation and Storage for Sensitive Data

August 22, 2019
UC Santa Cruz, Silicon Valley Campus
Santa Clara, CA

Purpose of Workshop

- ***Convene*** researchers from a variety of disciplines who are working with increasingly large, multi-dimensional datasets in the context of complex networks and computing infrastructure and who are often subject to an evolving landscape of protection requirements due to the involvement of human subjects or critical physical infrastructure
- ***Discuss*** new services and platforms being developed to provide secure research data management and computation, while balancing challenges of authentication, access control, and privacy protection

Who Attended?

- ***Approximately 57 registered participants***, including researchers, computer scientists, IT service providers, security architects and data scientists in academia and in the private sector from a variety of disciplines, including biology, medicine, public health, and the social sciences.

Program Committee

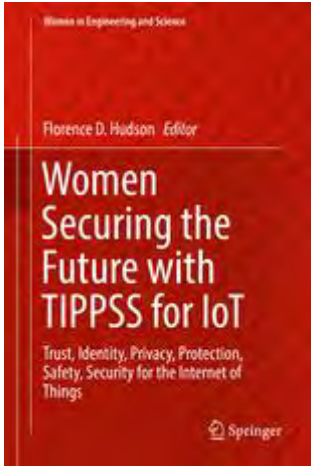
- Camille Crittenden, CITRIS and the Banatao Institute
- Tom DeFanti, UC San Diego, Qualcomm Institute, Calit2
- Chris Hoffman, UC Berkeley
- Jeff Weekley, UC Merced
- Mark Yashar, CITRIS, UC Berkeley

Workshop Agenda

9 am: Welcome -- David Rusting, Chief Information Security Officer, UCOP
9:15 am: Keynote Address -- *"TIPPSS (Trust, Identity, Privacy, Protection, Safety, Security) for Enabling and Securing our Increasingly Connected World,"* Florence Hudson, Founder and CEO at FDHint, LLC
10 am: Panel 1: Securing your Research Data: Perspectives from Domain Scientists -- Moderator: Jeffrey Weekley, Director of Cyberinfrastructure & Research Computing, UC Merced
11:15 am: Keynote Address -- *"Managing Complexity in a World of Surprise,"* David Alderson, Professor of Operations Research, Director of NPS Center for Infrastructure Defense, Naval Postgraduate School
1 pm: Panel 2: Securing your Infrastructure in a Changing Landscape: An Engineering Perspective -- Moderator: Jason Zurawski, Science Engagement Engineer, Energy Sciences Network (ESnet)
1:45 pm: Panel 3: Security and Privacy in Practice -- Moderator: Chris Hoffman, Associate Director of Research IT; Research, Teaching and Learning Services (RTL); UC Berkeley
2:30 pm: Closing Remarks -- Camille Crittenden, Executive Director, CITRIS and the Banatao Institute; co-founder of CITRIS Policy Lab and the Women in Technology Initiative at UC
3 pm: Adjourn

Keynote Address: "*TIPPSS (Trust, Identity, Privacy, Protection, Safety, Security) for Enabling and Securing our Increasingly Connected World*," Florence Hudson

- Provided a broad overview with a particular focus on trust, identity, privacy, protection, safety and security (the TIPPSS framework) around health data sharing, precision medicine, and medical devices, along with the human factors associated with cybersecurity.
- Specific areas of focus included:
  - Connectivity and the Internet of Things (IoT) in smart cities and campuses
  - Precision medicine will leverage large volumes and varieties of data. How do we protect the data, devices, and patients?
  - The creation and development of the TIPPSS framework



Panel 1: Securing your Research Data: Perspectives from Domain Scientists

Panel moderator: **Jeffrey Weekley**, Director of Cyber Infrastructure & Research Computing, UC Merced

Panel Members:

- **Peter Sudmant**, Assistant Professor, Dept of Integrative Biology, UC Berkeley
- **Matthew Renquist**, System Engineer, UC Davis Health
- **Rob Currie**, Chief Technology Officer, UC Santa Cruz Genomics Institute
- **Pavan Gupta**, Research and Cloud Computing Architect, UCSF Center for Digital Health Innovation

Panel 1: Insights from Domain Scientists

- Panelists brought experience from biology research, healthcare, and/or health-related fields, all requiring work with sensitive data
- Peter Sudmant described some of the real-world impacts on conducting research when various cybersecurity measures are applied and how patients can be directly affected.
- Rob Currie described cybersecurity measures and cyberinfrastructure used at the UCSC Genomics Institute.
- Matthew Renquist brought in his perspectives as a System Engineer with the Cancer Data Informatics Integration Initiative (CDI3) at UC Davis Health.
- Pavan Gupta then discussed his experience and research on how the Kubernetes open-source container-orchestration system could be implemented together and coupled with High Performance Computing tools securely.

Dockstore (dockstore.org), Rob Currie, UCSC

Keynote Address: “Managing Complexity in a World of Surprise,” David Alderson

Professor of Operations Research and director of the NPS Center for Infrastructure Defense at the Naval Postgraduate School in Monterey, CA, Alderson offered his perspectives and experiences in operations research with a focus on critical infrastructure systems.

Five key insights for critical digital services:

1. The Robust Yet Fragile (RYF) tradeoff means that it is not sufficient merely to add more and more technologies to “solve” the problems.
2. In a world of continuous deployments, we will never have complete knowledge of the system.
3. Use live experimentation to gain confidence in the system while “minimizing the blast radius”.
4. Traditional risk analysis is not sufficient for managing complexity in a world of surprise.
5. Big data analytics on their own are insufficient to avoid surprise.

Panel 2: Securing your Infrastructure in a Changing Landscape: An Engineering Perspective

Panel moderator: **Jason Zurawski**, Science Engagement Engineer, Energy Sciences Network (ESnet), Lawrence Berkeley National Laboratory

Panel Members:

- **Tolgay Kizilelma**, Chief Information Security Officer, UC Merced
- **Jim Basney**, Senior Research Scientist, National Center for Supercomputing Applications, University of Illinois, Urbana-Champaign
- **Igor Sfliligoi**, Lead Scientific Software Developer and Researcher, San Diego Supercomputer Center, UC San Diego
- **Aleatha Parker-Wood**, Machine Learning and Algorithmic Privacy Lead, Humu

Panel 2: Insights and Alarms from an Engineering Perspective

- Panelists discussed how changes in hardware, software, and network usage complicated security matters and what concerned them the most, with a focus on examples of specific risks and mitigations.
- They also discussed the current protections they were using in their work and to what extent those protections were successful, as well as future plans around security protections.
- Some expressed worries from a security perspective around the increasing prevalence of machine learning and blockchain technology.

Panel 2: Insights and Alarms from an Engineering Perspective

- Igor Sfliligoi, who has worked extensively with the PRP Nautilus Kubernetes cluster and the Open Science Grid (OSG) distributed High Throughput Computing (HTC) organization, along with other panel members, discussed the growing focus around software and systems such as Kubernetes, Jupyter, Singularity, HPC, cloud computing, and containerized systems and applications in general, and the associated security and privacy concerns around them.
- Tolgay Kizilelma along with the other panel members felt that the proliferation of multi-factor authentication has had a positive impact on the IT and cybersecurity community.

Panel 3: Security and Privacy in Practice

Panel Moderator: **Chris Hoffman**, Associate Director of Research IT; Research, Teaching and Learning Services (RTL) at UC Berkeley

Panel Members:

- **Steve Trush**, Research Fellow, Center for Long-Term Cybersecurity (CTLC), UC Berkeley; Deputy Director of Citizen Clinic at CTLC, UC Berkeley
- **Andrea Hesse**, Academic Divisional Computing Director, Information Technology Services, UC Santa Cruz
- **Sandeep Chandra**, Executive Director of Sherlock Cloud; Director of Health Cyberinfrastructure Division, San Diego Supercomputer Center
- **Elaine Sedenberg**, Privacy and Data Policy Manager at Facebook

Panel 3: Insights from Security and Privacy in Practice

- Panelists discussed their experiences developing, operating, and sustaining services for researchers working with sensitive data, opportunities for researchers and research institutions, as well as the broader organizational and societal issues involved.
- Key issues involved with protected data research:
  - Overhead to secure and manage data, account management, incident response, risk assessment
  - The business of running a secure hosting capability
  - Evolution of protected data services (e.g., making cloud services secure, easy to consume, and keeping costs down)
- Private sector companies also work with sensitive data that interest social science researchers (e.g., social media platform data), balance corporate priorities with regulatory requirements and responsibilities for protecting users' privacy.

Workshop Conclusion and Outcomes

- Through a post-workshop survey, attendees expressed that they gained insights into the campus level discussions around the complexity involved in securing the vast amounts of research data possessed by the UC system.
- Attendees also reported that they found both of the keynote presentations to be very informative and were very satisfied with the opportunities for professional networking during and after the workshop.
- The workshop agenda, speaker biographies, and presentation slide decks can be found on the PRP website: http://pacificresearchplatform.org/events/workshop-2019-srd/.

Workshop Key Takeaways

- Traditional risk analysis is not sufficient for managing complexity in a world of surprise
- Big data analytics on their own are insufficient to avoid surprise
- There is a growing focus in the cybersecurity community around securing technologies, software, and systems such as Kubernetes, Jupyter, Singularity, HPC, cloud computing, and containerized systems and applications in general.
- Organizations must engage researchers so they care about security, provide information & tools, financial support, etc.

Workshop Key Takeaways (cont'd.)

- Cybersecurity training should be part of academic curriculum.
- If grant awards require protected data, institutions should fund an operational capability for researchers to use and develop appropriate facilities.
- The need, and requirements, for data and privacy protection are only increasing, we need to adapt to this change.

Thank you!

Camille Crittenden, CITRIS and the Banatao Institute
ccrittenden@berkeley.edu | @camcritt

Jeff Weekley, UC Santa Cruz
jweekley@ucsc.edu | @jdweekley

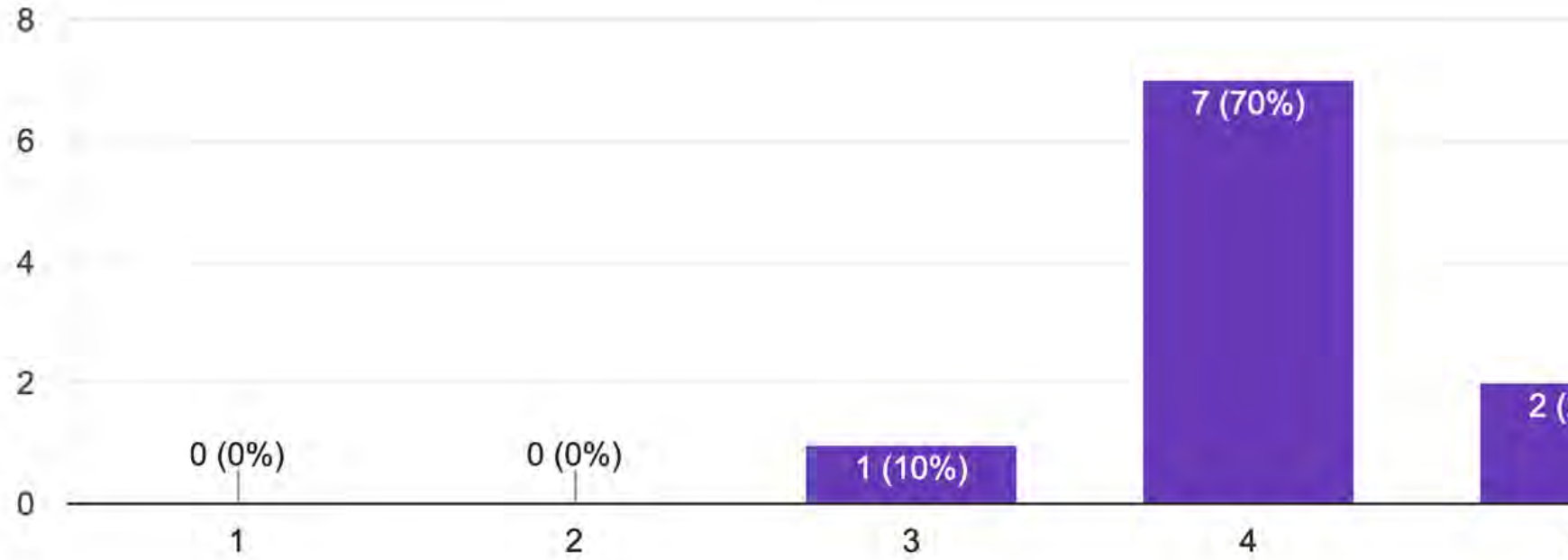
PACIFICRESEARCHPLATFORM.ORG

Additional Slides: Post-Workshop Survey Summary Results (10 Responses)

On a scale of 1 to 5 with 1 being "not satisfied" and 5 being "very satisfied"…

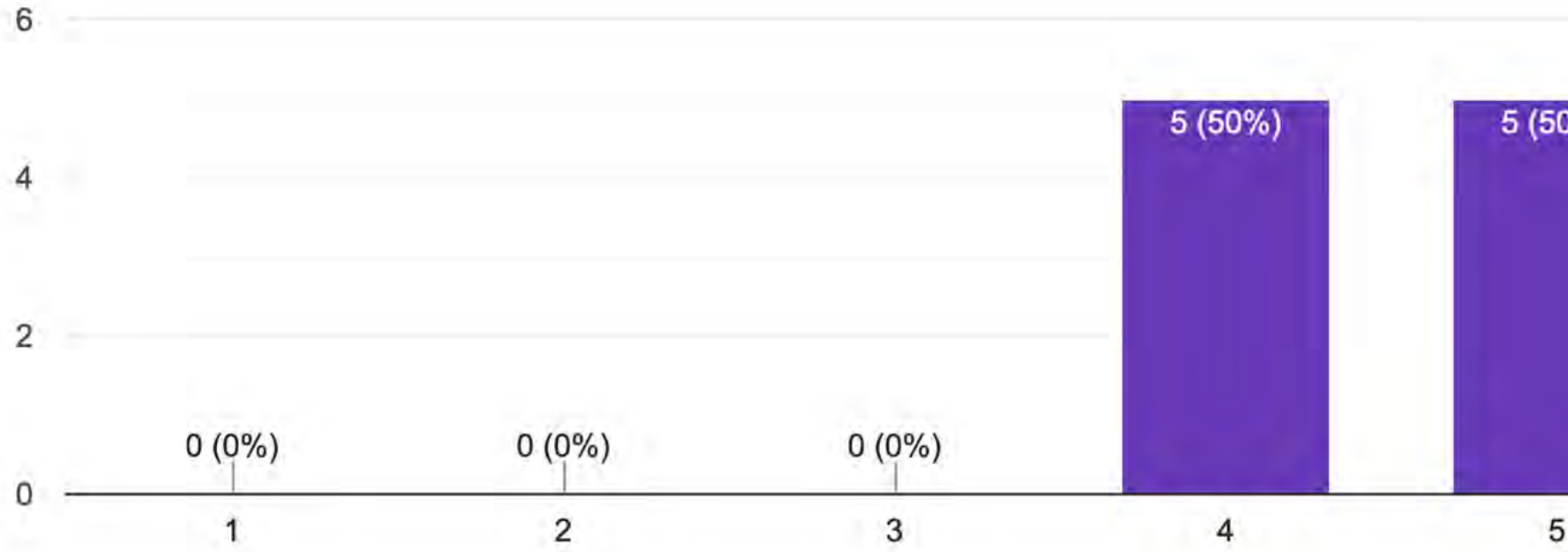
How satisfied were you with the overall workshop program?

10 responses



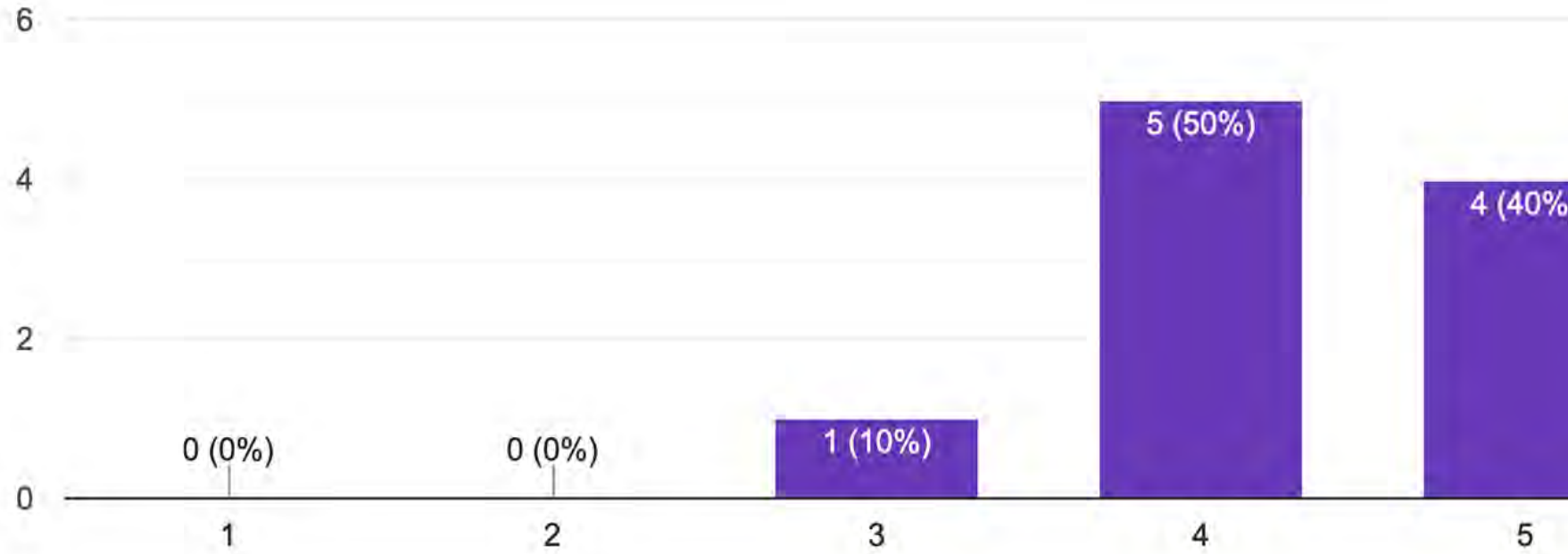
How satisfied were you with the keynote presentations?

10 responses



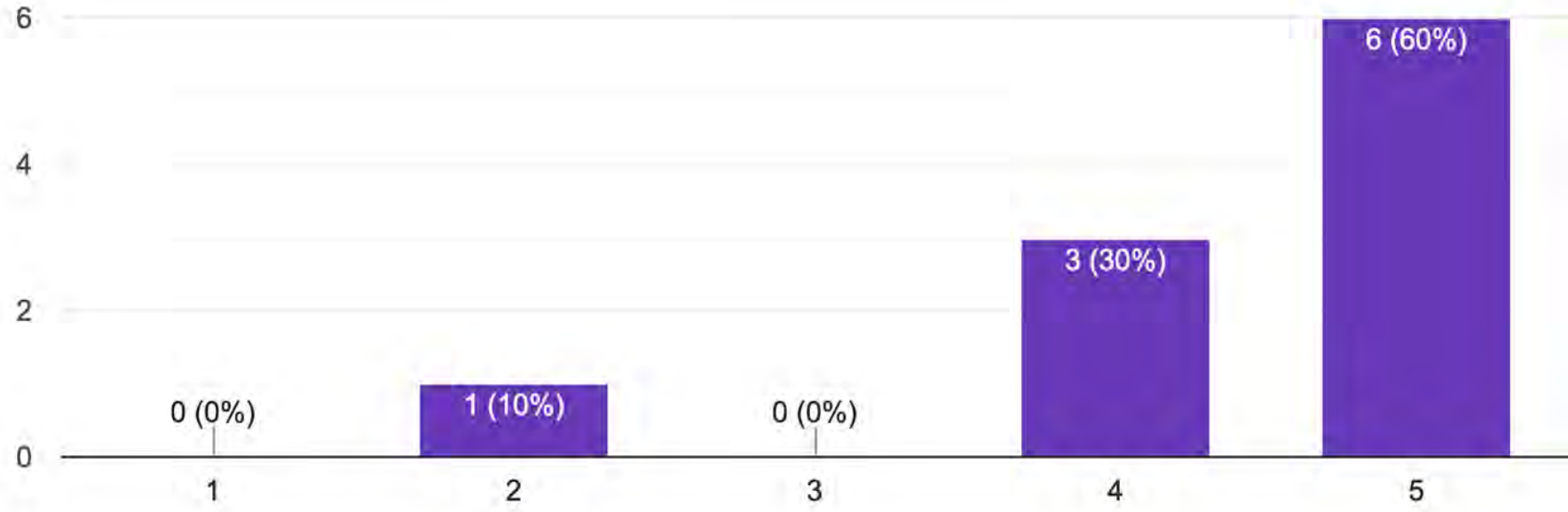
How satisfied were you with the panel themes and presentations?

10 responses



How satisfied were you with the opportunities for networking?

10 responses



How satisfied were you with the venue and logistics for the workshop?

10 responses