Cybersecurity Experimentation of the Future (CEF)

Jelena Mirkovic (USC/ISI), Srivatsan Ravi (USC/ISI)

Terry Benzel (USC/ISI), David Balenson (USC/ISI)

with input from many others in the community

22 February 2023

USC Viterbi School of Engineering Information Sciences Institute

This work was supported, in part, by the National Science Foundation under grant numbers 1925773, 1925616, 1925588, 1925564. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Why CEF?

- In the last three years we have seen:
  - A wide-reaching supply chain attack on government infrastructure - Solar Winds attack
  - A large ransomware attack on critical infrastructure - Colonial Pipeline
  - The largest cumulative DDoS attack to date - lasting 36 hours and generating total of 2.9 PB of traffic
  - Many privacy leaks, blunders and oversteps by technical companies
- Cybersecurity and privacy research are of critical importance for our daily lives, for our scientific progress and for critical infrastructure
- Reproducible experimentation is essential for research progress

Today's Research Landscape

- Our research is opportunistic:
  - Working on small, compartmentalized, simplified problems
  - Working with private datasets
  - Experimenting using resources in one's lab
  - Working in isolation from related work
- In the meantime:
  - Attacks are getting more sophisticated and coordinated
  - Attacks are getting stronger and more frequent
  - Attackers are specializing for certain types of attacks, and collaborating together

Today's Research Landscape

- Our research is opportunistic:
  - Working on small, compartmentalized, simplified problems
  - Working with private datasets
  - Experimenting using resources in one's lab *
  - Working in isolation from related work
- If we can improve reproducibility this would:
  - Increase sophistication of research solutions
  - Enable researchers to compare properly to related work
- To improve reproducibility we need:
  - Better research infrastructure
  - Better and more research artifacts ... that are easier to reuse

Too hard for one research group to work on complex problems. We need community resources and vertical research

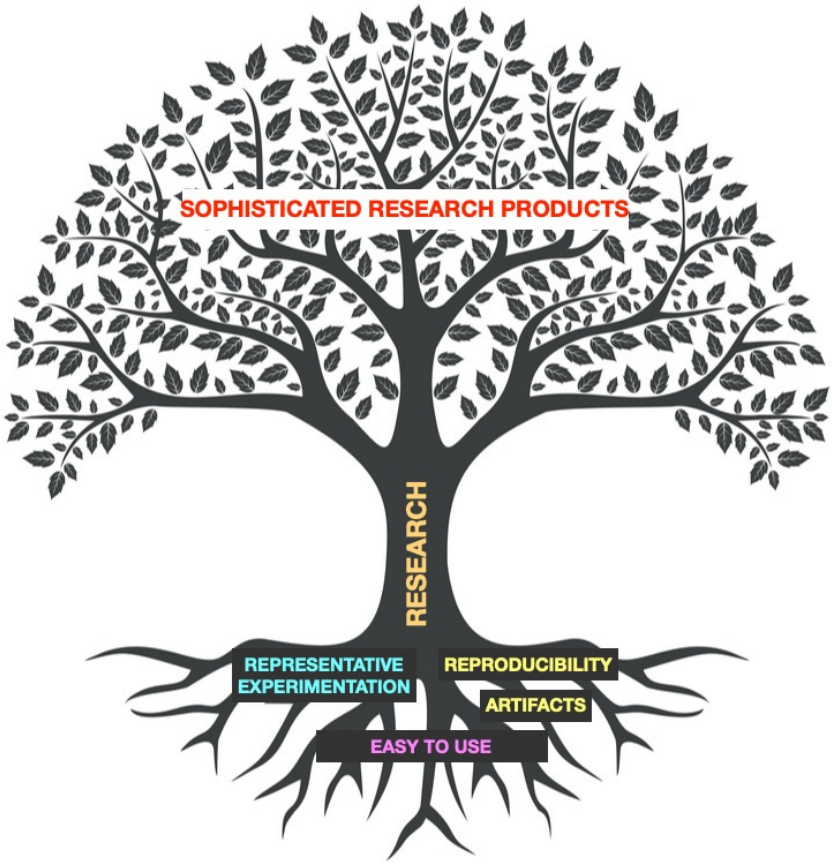
BUT

Low reproducibility

Take a Quick Survey

- https://bit.ly/LASER-exp

Future Directions



CEF 2014-18

- A series of study groups and community engagement workshops asking community input about future of cybersecurity experimentation:
  - Domains of applicability - multidisciplinary experimentation
  - Modeling the real world - human activity
  - Open interfaces for extensibility
  - Interconnected research infrastructure
  - Experiment design and instantiation - reusable designs for science-based hypothesis testing
  - Experiment execution and management
  - Instrumentation and experiment analysis
  - Meta-properties - usability and cultural changes

Good list, but did it change over time?
What exactly do we need and how to get there?

David Balenson, Laura Tinnel, and Terry Benzel, Cybersecurity Experimentation of the Future: Catalyzing A New Generation of Experimental Cybersecurity Research, Final Report, July 31, 2015.
https://cef.cyberexperimentation.org/application/files/2616/2160/7871/CEF_Final_Report_Bound_20150922.pdf

Experimentation: What is Missing?

- Most research is irreproducible
  - **CEF virtual workshop organized by USC/ISI in December 2022**
- Artifacts are shared in a way that makes them hard to reuse
  - **Artifacts virtual workshop organized by USC/ISI, University of Utah, UIUC and SRI International in September 2022**

CEF 2022 Workshop

- Around 30 participants from various cybersecurity and privacy research domains
  - Some also had experience in building research infrastructure (aka testbeds)
- We also circulated a survey via email to around 500 researchers
  - Received 58 responses
- Main questions:
  - What are experimentation needs?
  - What can testbeds do to meet them?
  - How to improve artifact sharing and reuse?

CEF 2022 Findings: Needs

- Common datasets and evaluation environments
  - So everyone works in the same setting, no rebuilding the world from scratch
  - Very research-domain dependent
- Modeling or including human users in experiments
  - So we can experiment with human factors

CEF 2022 Findings: Testbeds

- Representative experimentation environments
  - Same as experimentation need
- Amortize setup via reuse of packaged experiments
- User-friendly interfaces
  - Easy to learn
  - Easy to program/automate experimentation
- Ability to include third-party devices
  - No testbed will have all the hardware researchers need
- Variety of hardware and experimentation modes (e.g., simulation)
- Exposing testbed limitations to users

CEF 2022 Findings: Artifacts

- Incomplete artifacts
- Non-portable artifacts
- More artifact evaluation and research reproduction
  - Out of 96 security and privacy conferences only 6 have artifact evaluation
- Large storage for ML models
- Artifact packaging standards
- **Research infrastructure support for artifact packaging**

CEF 2022 Findings: Summary

- Community resources, representative environments and datasets
- … hosted on testbeds, which are easy to use and extensible
- ……with diverse hardware
- ……with ability to include humans in experiments
- ……with various experiment modes (e.g., simulation, emulation, measurement of real Internet)
- ……with help for packaging and sharing of artifacts

The CEF 2022 findings validate all findings from CEF 2014-2018

Take a Quick Survey

- https://bit.ly/LASER-art

Artifacts 2022 Workshop

- Around 32 participants from 18 organizations
  - Some also had experience chairing artifact evaluation committees
- We also circulated a survey via email to various mailing lists
  - Received 31 responses
- Main questions:
  - What are the challenges around artifact sharing and reuse?
  - Delve deeper into issues around:
    - Findability
    - Scope
    - Quality/usability
    - Evaluation
    - Community next steps

Artifacts 2022 Findings: Findability

- Artifacts shared in many different locations (e.g., Github, Zenodo, personal Web page, lab Web page)
  - Difficult to find
  - Difficult to establish relationship between artifacts
  - Catalogues would help here, but require a critical mass of users and artifacts (one example: https://hub.cyberexperimentation.org)
- Even when one finds an artifact, it is difficult to estimate how useful it is
  - Does it have relevant metadata? Hard to establish due to variable packaging
  - Did anyone else find it useful?
  - Is it maintained?

Artifacts 2022 Findings: Scope

- Artifacts are not only code and data
  - Also hypothesis, research methods, experiment design, preprocessing and postprocessing workflows, etc.
  - Experimentation environment may introduce biases, unbeknown to authors
- Authors are poorly trained to record and release these types of data
  - In some cases, too many details in a paper submission may decline chances of acceptance

Artifacts 2022 Findings: Quality

- Quality = usability
  - Good documentation, code is easy to run and understand
- Challenges for authors
  - Lots of effort to produce high-quality artifacts, maybe no one will use them
  - **No funding** – could we introduce easy to get, supplemental funding?
  - **Low impact on promotion, graduation progress or reputation**
  - No maintenance once lead student graduates
- Students need to be taught how to produce and package artifacts
  - Docker/VM, documentation, test cases

Artifacts 2022 Findings: Evaluation

- Lots of value for science, authors, venues and for use in education
- Authors may not feel that their artifact is ready
  - Lots of effort to make it ready, payoff is low
- Evaluators get almost no reward from evaluation
  - Hard to recruit skilled evaluators
- Main evaluation hurdle: special hardware and private datasets
- Should artifacts be required for publication? Or just encouraged?
  - If required, should they be evaluated?
  - Should we require them at submission time or at final version?

Artifacts 2022 Findings: Next Steps

- Standardization:
  - Need community standards around artifact packaging and quality
  - Need community guidelines/tutorials around sharing beyond code and datasets
  - Students need to learn best practices for sharing in grad school
- Incentives for authors and evaluators
  - Recognition, venues for artifacts only
- Build culture of sharing and reuse
- Provide funding for artifacts
  - E.g., supplements to current funded projects

Artifacts 2022 Findings: Summary

- We need high-quality artifacts that are also easy to find
- … need to educate and reward students to produce them
- … need to fund PIs to produce them
- … need to reward evaluators to identify quality artifacts
- … need to create research infrastructure that supports artifact packaging, sharing and reuse
- … need the community to build culture of sharing and reuse

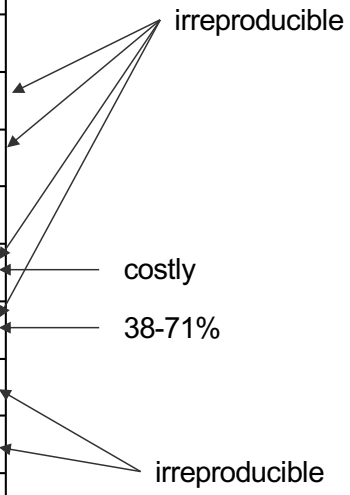
Conclusions

- We need more sophisticated cybersecurity and privacy research products
- … this rests on providing representative, easy-to-use experimental infrastructure and easy ways to share and reuse artifacts
- Our workshops produced a set of specific recommendations for the community, funding agencies, artifact authors and evaluators
- It will take a concerted effort of many to make progress
- … tutorials, classes, evaluation efforts at venues, funding supplements, reviewers asking for artifact release and comparison, etc.
- Progress may be non-linear, but we should persist

Paper Survey – Experimentation Practices

- Surveyed 704 papers from top four cybersecurity conferences in 2022

Venue	USENIX Sec	Oakland	NDSS	ACM CCS
Papers	257	146	83	218
Papers w experiments	252	130	83	193
Measurement	40	13	3	13
Survey	27	8	3	6
Dataset analysis	14	7	2	1
Cloud	17	6	5	21
Own institution	98	72	47	137
Binary analysis	22	14	11	3
Special HW	43	21	12	24
Simulator	9	2	5	2
Testbed	3	1	1	0
General compute	87	39	27	78



Experimentation: What is Missing?

- Most research is irreproducible
  - Instead of using public testbeds researchers are using their own computers or paying for clouds
  - Around 35% of experiments could be done using general compute nodes, present in most public testbeds
- Artifacts are shared in a way that makes them hard to reuse
  - Hard to find
  - Inconsistent packaging (zip files, Github repos, Web pages)
  - May lack important information
  - May have hard-coded data and implicit assumptions
  - May have missing dependencies

How can we do better in the future?