Analyzing Cyber Security Research Practices through a Meta-Research Framework

Victor Le Pochat, Wouter Joosen

CSET 2023, 7 August 2023

Analyzing Cyber Security Research Practices through a Meta-Research Framework

Analyzing cyber security research practices

- Cyber security research should be **valid** and **sound**
  - Use appropriate methods, evaluate and communicate well, ...
  - Ensure reliable results, correct findings and well-justified claims
  - Accurately reflect real-world security & propose effective solutions
- Understanding *how* we do research can help us *improve* it

Analyzing Cyber Security Research Practices through a Meta-Research Framework

Meta-research

“Do research on research”

- Critically **evaluate** research practices
- Understand if research is **sound** and **reliable**
- Assess presence/mitigation of research **biases**

Analyzing Cyber Security Research Practices through a Meta-Research Framework

Ioannidis et al. [Ioa15] introduced a framework for categorizing **meta-research** work

› *Methods*	performing research as best as possible
› *Reporting*	communicating research well
› *Reproducibility*	verifying research by reproducing it
› *Evaluation*	fairly evaluating research by peer review
› *Incentives*	rewarding research correctly and fairly

Analyzing Cyber Security Research Practices through a Meta-Research Framework

We **categorize** cyber security meta-research work

“Do meta-research on meta-research”

- **Goal**: gain a better understanding of our community’s efforts to examine its own research practices
- **Process**: apply the framework by Ioannidis et al.
- **Result:** characterize main areas of meta-research work
- Encourage our community to continue self-reflection

Categorization

Methods

Reporting

Reproducibility

Evaluation

Incentives

Methods

Reporting

Reproducibility

Evaluation

Incentives

Methods: *performing*

Goal: Conducting research using the best scientific methods & practices (available)

Risk: Experiments and results are not truly representative or accurate

Interest: high

Methods: *performing*

- Best practices and pitfalls
  - Common in many domains of cyber security research

    Malware, machine learning, hardware, systems, social networks, …
  - Correct and open data sets, proper metrics and benchmarks, …
  - Possibly “flawed” prior work as examples
  - Serve as a reference for future studies

Methods: *performing*

- Qualitative methods
  - Usually for usable security and privacy
  - Special care/scrutiny to show validity of research methods
    - Rest of security community: unfamiliar
  - Best practices and guidelines, specific to qualitative methods
    - But not always followed [Gro20,Kau21]

Methods: *performing*

- Ethical considerations
  - Existing frameworks for ethical review (e.g., IRB) may be unadapted
  - Community has to set own ethical standards (+ provide guidelines)
    - Security: Menlo Report
    - Ethics increasingly enforced at top-tier conferences
  - Controversial studies serve as use cases for lessons learned

Methods

Reporting

Reproducibility

Evaluation

Incentives

Reporting: *communicating*

Goal: Reaching the intended audience(s) with research results relevant to them

Risk: Results are misinterpreted or misrepresented

Interest: medium

Reporting: *communicating*

- Publication bias
  - Under or overrepresented research
    - e.g., omission of negative results (shown in security user studies [Gro20])
    - e.g., more attack than defense papers?
- Preregistration
  - Stabilize research questions, hypotheses, methods, analyses, ... *before* actual experiments take place
  - Very uncommon in security and privacy research
    - Due to exploratory or vulnerability-driven nature of studies?

Methods

Reporting

Reproducibility

Evaluation

Incentives

Reproducibility: *verifying*

Goal: Repeating a study to confirm its results and increase the likelihood that its hypothesis is correct

Risk: Failing to repeat a study puts validity of its results into question → “replication crisis” (?)

Interest: high

Reproducibility: *verifying*

- Artifacts
  - Sharing data sets and tools
    - Artifact evaluation (badges)
    - Still often fail to meet replicability criteria [Dem22]

Methods

Reporting

Reproducibility

Evaluation

Incentives

Evaluation: *evaluating*

Goal: Judging the quality of a research paper
to maintain the integrity of science

Risk: Subjectivity could lead to published subpar papers
and unpublished state-of-the-art-advancing papers

Interest: medium

Evaluation: *evaluating*

- Peer review
  - Top-tier security conferences [Son22]
    - *Novelty* as only shared evaluation metric
    - Various reasons to reject (“toxic culture of rejection”? [Lee22])
    - Sense of ‘randomness’

Evaluation: *evaluating*

- Peer review
  - Trend towards journal-style model (i.e., revisions)
  - Good reviewing practices encouraged (meta-reviews, awards, ...)

  - A lot of trials, but also a lot of reversals?

Methods

Reporting

Reproducibility

Evaluation

Incentives

Incentives: *rewarding*

Goal: Evaluating the quality, value, and impact of research and providing the right incentives and support

Risk: Incentivizing “wrong” research (practices), improperly supporting “good” research

Interest: *medium-low*

Incentives: *rewarding*

- Rankings
  - Conference (tiers)
    - More restrictive = more prestigious
    - “Underappreciated” research? (e.g., replication studies)
  - Researchers, institutions
    - Criticism: biased or non-representative of quality

Discussion and conclusion

More meta-research work is being published

- Strong focus on improving ***methods***
  - Best practices, analyzing data collection tools, data sets, ...
  - Lack of central repository may make awareness & adoption difficult
  - Enforcement: left as a task for peer review?
- ***Other categories***: less work, but more clarity
  - Enforced or encouraged explicitly, with noticeable evolutions
    e.g., ethical considerations, artifact badges, stricter peer review
  - Less (academic) reflection?

Meta-research is a collective community effort

- Venues like CSET support discussion of research practices
- Research communities can **learn from each other**
  - Meta-reviews: gone in Internet measurement, back in security?
  - *Introspectively: framework from biology can be reused in security*
  - Some concerns are common to all fields (e.g., incentivization)
- But all communities have their **own accents**
  - Badging as artifact encouragement; lack of preregistration

Cyber security meta-research

contributes to more reliable and trustworthy cyber security research

and therefore helps to improve cyber security itself

Analyzing Cyber Security Research Practices through a Meta-Research Framework

victor.lepochat@kuleuven.be @VictorLePochat

https://lepoch.at

References

- [Her17] Cormac Herley and P.C. Van Oorschot. 2017. SoK: Science, Security and the Elusive Goal of Security as a Scientific Pursuit. In 2017 IEEE Symposium on Security and Privacy. 99–120. https://doi.org/10.1109/SP.2017.38
- [Spr17] Jonathan M. Spring, Tyler Moore, and David Pym. 2017. Practicing a Science of Security: A Philosophy of Science Perspective. In 2017 New Security Paradigms Workshop. 1–18. https://doi.org/10.1145/3171533.3171540
- [Ioa15] John P. A. Ioannidis, Daniele Fanelli, Debbie Drake Dunne, and Steven N. Goodman. 2015. Meta-research: Evaluation and Improvement of Research Methods and Practices. PLOS Biology 13, 10, Article e1002264 (October 2015). https://doi.org/10.1371/journal.pbio.1002264
- [Bas16] Aniqua Baset and Tamara Denning. 2019. A Data-Driven Reflection on 36 Years of Security and Privacy Research. In 12th USENIX Conference on Cyber Security Experimentation and Test.
- [Gro20] Thomas Groß. 2020. Statistical Reliability of 10 Years of Cyber Security User Studies. In 10th International Workshop on Socio-Technical Aspects in Security and Trust. 171–190. https://doi.org/10.1007/978-3-030-79318-0_10
- [Kau21] Mannat Kaur, Michel van Eeten, Marijn Janssen, Kevin Borgolte, and Tobias Fiebig. 2021. Human Factors in Security Research: Lessons Learned from 2008-2018. https://doi.org/10.48550/arxiv.2103.13287 arXiv:2103.13287
- [Nar23] Arvind Narayanan and Kevin Lee. 2023. Security Policy Audits: Why and How. IEEE Security & Privacy 21, 2 (2023), 77–81. https://doi.org/10.1109/MSEC.2023.3236540
- [Bal22] David Balenson, Terry Benzel, Eric Eide, David Emmerich, David Johnson, Jelena Mirkovic, and Laura Tinnel. 2022. Toward Findable, Accessible, Interoperable, and Reusable Cybersecurity Artifacts. In 15th Workshop on Cyber Security Experimentation and Test. 65–70. https://doi.org/10.1145/3546096.3546104

References

- [Pen20] Jan Pennekamp, Erik Buchholz, Markus Dahlmanns, Ike Kunze, Stefan Braun, Eric Wagner, Matthias Brockmann, Klaus Wehrle, and Martin Henze. 2020. Collaboration is not Evil: A Systematic Look at Security Research for Industrial Use. In 2020 Learning from Authoritative Security Experiment Results Workshop. 16 pages. https://doi.org/10.14722/laser-acsac.2020.23088
- [Bou22] Nicholas Boucher and Ross Anderson. 2022. Talking Trojan: Analyzing an Industry-Wide Disclosure. In 2022 ACM Workshop on Software Supply Chain Offensive Research and Ecosystem Defenses. 83–92. https://doi.org/10.1145/3560835.3564555
- [Son22] Ananta Soneji, Faris Bugra Kokulu, Carlos Rubio-Medrano, Tiffany Bao, Ruoyu Wang, Yan Shoshitaishvili, and Adam Doupé. 2022. “Flawed, but like democracy we don’t have a better system”: The Experts’ Insights on the Peer Review Process of Evaluating Security Papers. In 2022 IEEE Symposium on Security and Privacy. 1845–1862. https://doi.org/10.1109/SP46214.2022.9833581
- [Dem22] Nurullah Demir, Matteo Große-Kampmann, Tobias Urban, Christian Wressnegger, Thorsten Holz, and Norbert Pohlmann. 2022. Reproducibility and Replicability of Web Measurement Studies. In ACM Web Conference 2022. 533–544. https://doi.org/10.1145/3485447.3512214
- [Lee23] Edward Lee. 2022. The Toxic Culture of Rejection in Computer Science. ACM SIGBED. https://sigbed.org/2022/08/22/the-toxic-culture-of-rejection-in-computer-science/
- [Ort22] Anna-Marie Ortloff, Matthias Fassl, Alexander Ponticello, Florin Martius, Anne Mertens, Katharina Krombholz, and Matthew Smith. 2023. Different Researchers, Different Results? Analyzing the Influence of Researcher Experience and Data Type During Qualitative Analysis of an Interview and Survey Study on Security Advice. In 2023 CHI Conference on Human Factors in Computing Systems. Article 864, 21 pages. https://doi.org/10.1145/3544548.3580766