DOI: 10.1145/3651965

BY VICTOR LE POCHAT

Reflecting on Research Practices

Encouraging the security research community to engage in meta-research.

There are growing concerns about the quality, reliability, and repeatability of research results, across the broader computing community and more specifically in cyber security as well. In this *Communications* Security column, I highlight the need and value of studying research practices themselves—also called *meta-research*—to address those concerns. Meta-research contributes to improving the research process and supporting researchers in preserving the soundness and validity of their research. I will show how the security research community has already made good progress in engaging in such introspection, especially regarding its methods. Nevertheless, we should continue to expand these efforts to the whole research cycle, to further improve our research practices and ensure we can trust scientific findings to be reliable and reproducible.

Background: The Science of Security

Across all scientific disciplines, we increasingly see researchers questioning whether current scientific practices produce trustworthy research findings. For example, the 'replication crisis' emphasizes how studies may be difficult or even impossible to reproduce, which threatens the validity of their results. This has even led to statements that "most published research findings are false."[2] The security research community is not immune to such concerns: a recent study by Soneji et al. gave insight into potential shortcomings of the peer review process at top-tier security conferences,[5] such as subjective evaluation metrics and perceived randomness in decisions.

In cyber security, the *science of security* encapsulated efforts to define, analyze, and promote better scientific practices,[1,6] starting to take shape approximately 15 years ago, with the ultimate intent to build more intrinsically secure systems. A decade later, a community debate ensued concerning whether these efforts had been successful, even questioning whether security can be considered a "science." The critics saw two major flaws: current security research is not conducted scientifically, failing to adopt practices that the 'rest of science' has implemented over time; and security inherently does not lend itself to the scientific method, with unfalsifiable security claims along with a failure to transfer academic models into real-world outcomes.[1] The proponents refute these obstacles, finding that they rely on outdated views on science. Based on a more modern view, they find the research community currently already engages in good scientific practice.[6]

This philosophical debate on what science *is*, and how security research may or may not fit that definition, distracts us from concrete efforts to improve security research practices. Regardless of where one may stand in this debate, we should always strive to be rigorous in our approach to research. In fact, this appears to be a matter where the critics and proponents of the 'science of security' agree: the application of good scientific practices is a way to move the security field forward.

Meta-Research

The field of meta-research is highly relevant to this discussion. Its goal is to critically evaluate current research methods and practices, examine biases threatening the soundness and reliability of research, and develop new best practices for conducting research.3 I argue that giving serious consideration to our security research practices through meta-research is a crucial step toward ensuring valid and sound research, and deserves more attention from the community. Without such 'research on research', we cannot be confident that we can fully trust research findings.

Other communities have more experience with meta-research. In biology, Ioannidis et al.[3] introduced a categorization for meta-research areas, which serves as a benchmark for holistically evaluating research practices. They identified five different phases of the research cycle, covering the whole scientific process:

- *methods*, or how studies are performed, with biases from flawed datasets or methods, sampling biases, to inappropriate statistical analyses;
- *reporting*, or how research is communicated, with biases due to misrepresentation or misinterpretation of results;
- *reproducibility*, or how research is verified, with biases from undiscovered errors or one-off observations;
- *evaluation*, primarily peer review, with biases such as favoring positive results or randomness in paper acceptance; and
- *incentives*, or how research is rewarded, with biases such as overreliance on paper metrics (citations) or subjective funding criteria.

To gain a better understanding of the security research community's ongoing efforts to examine and improve its own research practices, we applied this framework to security meta-research work,[4] to identify the main areas and gaps, and outline current practices, priorities, or open questions on how the community conducts security research. In this column, I highlight the main insights into current security research practices from this study and conclude with specific recommendations.

Insights into Current Security Research Practices

In our study, we found that not all categories of security meta-research receive the same amount or kind of attention from the community. Researchers put a strong focus on scrutinizing methods, from a (correct) belief that sound methods are fundamental to obtaining reliable findings. Researchers across a large variety of security domains have identified significant pitfalls in existing methods and developed best practices going forward. Interestingly, this is sometimes out of necessity. For example, the usable security and privacy community takes special care to show the validity of its—usually qualitative—methods. This is perhaps needed to 'convince' the rest of the community, who might be unfamiliar with these methods, that such methods are valid and appropriate. We also see a growing body of critical analyses of commonly used data collection tools or data sets. Fortunately, this does not end at only identifying these flaws, as the community is actively proposing improved tools and datasets to enable other researchers to make their own research more sound.

While we identified in our study that best or worst practices have been established, that does not mean researchers will automatically apply or avoid them, respectively. Even top-tier security work regularly succumbs to pitfalls. Understandably, in a rapidly evolving and growing community, it can be hard to keep up with the latest insights into state-of-the-art practices. It is not enough to lean on what prior work has done, as that work may very well rely on outdated practices. In addition, the time span between selecting methods for a research project and publishing its results may inherently delay a broad adoption of best practices. Equally, we cannot expect peer review to always enforce

the use of state-of-the-art methods, as it is untenable to ask from individual reviewers that they are fully aware of the constantly evolving best practices, especially if reviewers are not familiar with a given (sub)field. We would recommend that the community develops a central repository to serve as a collective, up to date, and easily referenceable resource of the best practices for which there is community consensus. This would support researchers in discovering and applying those methods, as well as reviewers in properly assessing whether papers apply those methods.

Compared to methods, we found in our study that security meta-research literature has less coverage for the other phases of the research cycle. Instead, the research community more explicitly encourages or enforces desirable practices. For example, security conferences have begun to require that papers thoroughly reflect on ethical considerations, after controversial cases emerged. Notably, the community had to develop the expectations and standards regarding ethics itself, as existing frameworks (for example, Institutional Review Boards) proved to be inadequate. Similarly, the introduction of artifact evaluation processes and their encouragement through badging supports reproducibility. The peer review process also continuously evolves, increasing accountability through, for example, public reviews and incorporating (journal-style) revisions, although some conferences have already started to backtrack on these developments.

These changes play an integral role in aspiring to higher scientific rigor and objectivity, and making these practices more explicit provides clarity as to what the community expects. Nevertheless, measures are often introduced and later reverted only because of concerns or 'feelings', while they are lacking (academic) reflection and actual empirical evidence. Without understanding whether such changes are effective at their intended goals, we as a community may waste time and effort, especially if we unthinkingly retry measures that have been shown to be ineffective. As an illustration, ACM IMC, a top Internet measurement conference, stopped publishing meta-reviews (public summaries of a paper's reviews) in 2014 because a community survey found no clear benefits, yet IEEE S&P, a top-tier security conference, introduced them 10 years later. Of course, the challenge is that we need a good set of metrics to empirically evaluate these changes, as well as a baseline to compare against. We should not shy away from trying new ways of improving the research process, but should establish how we will measure their effectiveness, to substantiate that changes are necessary and useful, and objectively justify reverting actions that turn out to not work.

One solution is for our community to look more at the experiences of other research communities and learn from them. We did so ourselves in our study, reusing a framework for analyzing research practices from the biology community, which shows the benefit of knowledge sharing across disciplines. However, we still need researchers who have sufficient domain knowledge to execute meta-research in the security field, as the topics are community-specific, with their own focus and prevalence. For example, the security community primarily encourages reproducibility through artifact badging, while other communities see reproducibility more as a topic for academic study or even something that is always expected or required. Of course, I extend this call for reflection and interaction to the entire computing community, who should similarly engage in meta-research as a means to analyze and improve their research practices.

Conclusion

From our study of (and contribution to) the growing body of security meta-research work, I derive several recommendations for encouraging further reflection on scientific practices. There are still gaps left to be filled, notably when it comes to studying research processes beyond the core methods of a paper, but also establishing more explicit and discoverable guidance for those methods. I see a great opportunity to continue the rigorous, critical, and academic analysis that we are accustomed to for scrutinizing and improving methods. This should be expanded to other phases of the research cycle such as reproducibility or peer review, to improve how we understand their impact on our research. There is certainly a lot of value in pursuing research in this direction of improving the soundness and validity of how we do science.

I strongly believe that this should be a larger, collective community effort. Several venues welcome discussions of the processes and practices that form research, and I encourage everyone to submit meta-research papers to these venues and follow the work that is published there. Within

computer security, these include the *Cyber Security Experimentation and Test* (CSET) and *Learning from Authoritative Security Experiment Results* (LASER) workshops. I find sound and valid research practices to be of utmost importance to the reliability of research findings, so I would also strongly recommend increasing coverage of these topics at the main security conferences and journals. This would allow the field of meta-research to grow and reach larger audiences, proving that our community truly values this work.

References

1. Herley, C. and Van Oorschot, P.C. Sok: Science, security and the elusive goal of security as a scientific pursuit. In *Proceedings of the 2017 IEEE Symp. on Security and Privacy*, 2017; doi: 10.1109/SP.2017.38
2. Ioannidis, J. Why most published research findings are false. *PLoS Medicine 2*, 8 (Aug. 2005); 10.1371/journal.pmed.0020124
3. Ioannidis, J. et al. Meta-research: Evaluation and improvement of research methods and practices. *PLOS Biology 13*, 10 (Oct. 2015); doi: 10.1371/journal.pbio.1002264
4. Le PochatV. and Joosen, W. Analyzing cyber security research practices through a meta-research framework. In *Proceedings of the 2023 Cyber Security Experimentation and Test Workshop (CSET '23)*, 2023; doi: 10.1145/3607505.3607523
5. Soneji, A. et al. Flawed, but like democracy we don't have a better system: The experts' insights on the peer review process of evaluating security papers. In *Proceedings of the 2022 IEEE Symp. on Security and Privacy*, 2022; doi: 10.1109/SP46214.2022.9833581
6. Spring, J.M. et al. Practicing a science of security: A philosophy of science perspective. In *Proceedings of the 2017 New Security Paradigms Workshop*, 2017; doi: 10.1145/3171533.3171540

Victor Le Pochat (victor.lepochat@kuleuven.be) is a postdoctoral researcher at the DistriNet research unit of the Department of Computer Science at KU Leuven, Belgium.

The author thanks the participants of the 16th Cyber Security Experimentation and Test Workshop, and the organizers David Balenson and Terry Benzel. We also thank Wouter Joosen, Lieven Desmet, Michel van Eeten, Maciej Korczyński, Frank Piessens, and Katrien Verbert for their comments while developing this column.

This research is partially funded by the Research Fund KU Leuven, and by the Cybersecurity Research Program Flanders.