Initial Thoughts on Cybersecurity And Reproducibility

Ewa Deelman[1], Victoria Stodden[2], Michela Taufer[3], and Von Welch[4]

[1]Information Sciences Institute [2]University of Illinois at Urbana-Champaign
[3]The University of Tennessee Knoxville [4]Indiana University

CENTER FOR APPLIED CYBERSECURITY RESEARCH
PERVASIVE TECHNOLOGY INSTITUTE

A Little Background

- I lead the NSF Cybersecurity Center of Excellence: trustedci.org
- TrustedCI's mission in a nutshell: (1) Understand how cybersecurity supports science and (2) help community implement cybersecurity.
- What cybersecurity supporting science means: Productivity, Trustworthiness, Reproducibility

A Conversation Starter

1. A computer system without cybersecurity is not predictable, and hence would unable to provide reproducibility.
2. However, there exists a number of challenges in how cybersecurity contributes to and challenges reproducibility.
3. Our paper attempts to enumerate these challenges and start a conversation between cybersecurity and reproducibility researchers and professionals.

We Believe This is Unexplored Territory

We are aware of work on the reproducibility of cybersecurity experiments (e.g. [1]).

We are not aware of any research as to the impact of cybersecurity, both positive and negative, on reproducibility.

This paper is about the latter.

Cybersecurity Definition

We broadly include:

- Preventing malicious intrusions.
- Preventing denial of use
- Preventing data alteration, intended or not.
- Preventing privacy violations.
- Confidentiality of source code and other research artifacts.

Reproducibility Definition

We define reproducibility in the computational sense: providing digital scholarly objects associated with the computational findings that would allow a reader to understand and regenerate the results. This includes any data, codes or scripts, inputs, and other relevant information, and made available in an open way if possible. [2, 3].

Reproducibility and Cybersecurity Challenges

Impact of Unauthorized Access on Reproducibility

- Unauthorized access -> loss of confidence that the computer system is behaving as it is intended.
- Can be restored to some extend through forensics and investigation.
- How does this loss of confidence impact reproducibility?

Impact of Patching on Reproducibility

- Patching: fixing a cybersecurity vulnerability on a operating system.
- Ideally a patch doesn't otherwise impact system behavior or performance.
- In the real world, this ideal doesn't hold.
  - E.g. Spectre and Meltdown patches had significant impacts on system performance [4].
- When do such changes impact reproducibility?

Impact of Imperfect Data Integrity on Reproducibility

- Data integrity errors may be caused maliciously or by IT failures.
- With larger data sizes, changes of IT failures is growing.
- Different science domains seem to have different tolerances.
- Can we quantify when data integrity errors are harmful to reproducibility?

Confidentiality of Data and Software

- Reproducibility relies on availability of software and data used in research.
- What if data has privacy issues? Or software is not open source?
- When and how can reproducibility accept confidential research artifacts?

Cybersecurity as an Ethical Issue

- Do cybersecurity failures lead to a lack of confidence in a computer system by the public and other researchers?
- Does that lack of confidence translate into a lack of confidence in the scientific results which were generated by using that computer system?
- Is this a motivating need reproducibility needs to address?

Trading off Reproducibility and Productivity

- Cybersecurity is not free, for example:
  - Implementation costs.
  - Performance overhead.
  - System complexity/usability.
- What is the appropriate trade-off point for reproducibility?

Closing Thoughts

1. Hopefully convinced you there are some interesting challenges at the intersection of cybersecurity and reproducibility.
2. A goal of cybersecurity for science should be reproducibility.
3. The authors will continue to refine and research the issues described and welcome collaboration.

Acknowledgments

Thank you to P-RECS reviewers for their feedback.

This work is supported by the National Science Foundation under Grants 1547272, 1642070, 1642053, 1642090, 1813537, and 1823385. The views expressed do not necessarily reflect the views of the National Science Foundation or any other organization.

Paper citation: Ewa Deelman, Victoria Stodden, Michela Taufer, and Von Welch. 2019. Initial Thoughts on Cybersecurity And Reproducibility. In 2nd International Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS'19), June 24, 2019, Phoenix, AZ, USA. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3322790.3330593

Presentation References

[1] Sean Peisert and Matt Bishop. 2007. How to Design Computer Security Experiments. In Fifth World Conference on Information Security Education, Lynn Futcher and Ronald Dodge (Eds.). Springer US, Boston, MA, 141–148.

[2] Victoria Stodden. [n. d.]. The Legal Framework for Reproducible Scientific Research. IEEE Computing in Science and Engineering 11, 1 ([n. d.]), 35–40. https://doi.org/10.1109/MCSE.2009.19

[3] Victoria Stodden. 2013. Resolving Irreproducibility in Empirical and Com- putational Research Âń IMS Bulletin. http://bulletin.imstat.org/2013/11/ resolving- irreproducibility- in- empirical- and- computational- research/. Ac- cessed: 2019-4-6.

[4] Peter Bright. 2018. Here's how, and why, the Spectre and Meltdown patches will hurt performance. https://arstechnica.com/gadgets/2018/01/ heres-how-and-why-the-spectre-and-meltdown-patches-will-hurt-performance/. Accessed: 2019-4-8.