Towards an Ecosystem for Reproducible Research in Computer Networking

Quirin Scheitle
Technical University of Munich
scheitle@net.in.tum.de

Matthias Wählisch
Freie Universität Berlin
m.waehlisch@fu-berlin.de

Oliver Gasser
Technical University of Munich
gasser@net.in.tum.de

Thomas C. Schmidt
HAW Hamburg
t.schmidt@haw-hamburg.de

Georg Carle
Technical University of Munich
carle@net.in.tum.de

ABSTRACT

Reproducibility is key to rigorous scientific progress. However, many publications in the computer networks community lack support for reproducibility. In this paper, we argue that the lack is mainly rooted in the additional effort that authors need to spend, without expecting sufficient benefits. Based on our experience in both authoring reproducible research and reproducing publications, we propose an ecosystem that incentivizes authors and reproducers to invest additional effort. This ecosystem consists of various building blocks, which can be combined into venue-specific profiles. A key building block is the Reproducibility Challenge, which we suggest to co-locate with the annual SIGCOMM conference to leverage reproducibility research in practice.

CCS CONCEPTS

• **General and reference** → **Experimentation**; **Validation**; • **Networks** → **Network performance evaluation**; • **Social and professional topics** → **Computing education**;

KEYWORDS

Reproducibility, Networks, Reproducibility Challenge

ACM Reference format:
Quirin Scheitle, Matthias Wählisch, Oliver Gasser, Thomas C. Schmidt, and Georg Carle. 2017. Towards an Ecosystem for Reproducible Research in Computer Networking. In *Proceedings of Reproducibility '17, Los Angeles, CA, USA, August 25, 2017,* 4 pages.
https://doi.org/10.1145/3097766.3097768

Reproducibility '17, August 25, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5060-0/17/08…$15.00
https://doi.org/10.1145/3097766.3097768

1 INTRODUCTION

Scientific results should be repeatable, replicable, and reproducible as defined by the ACM [3]. The goal of reproducible research in computer networks has already been discussed in a similar workshop at ACM SIGCOMM 2003 [4], without visible take-up of discussed suggestions. Only some papers in the computer networking community (*e.g.,* [7, 8]) include sufficient details to reproduce results presented in the publication. Most papers do not provide any artifacts, which also rules out replication.

There are multiple reasons that hinder reproducible research. First of all, a lack of common terminology and expectations may result into submissions that claim reproducibility but only make small parts of a publication actually reproducible, if at all. Another problem is limited access to data, even if a measurement methodology to gather the data is well explained. Furthermore, publication of the required details also comes with a risk to authors—external scrutiny may uncover flaws, possibly questioning insights of the paper. On the other hand, a very detailed description may distract the reading flow and thus reduce chances of acceptance. Most of these aspects can be solved, but require additional effort by the authors. And it seems that authors do not experience sufficient benefits which balance the effort.

A simple solution could be a mandatory reproducibility check as part of the reviewing process. Similar to the IETF that aims for running code of protocol specifications, the scientific community should aim for at least one practical reproduction of results to assess artifacts and results properly. However, reviewers already face a high review load, and they might not have the capabilities to carefully assess a practical implementation, even though they are experts in their field. Also, careful reproduction often requires interaction with authors, at least during the first round of reproduction, which conflicts with a blind review process. Besides authors and official reviewers, independent reproducers lack incentives as well. As venues, whether journal, conference, or workshop, typically focus on novel ideas, reproduction papers are rarely accepted for publication, even if they significantly extend knowledge through refined methodology or more heterogeneous measurements.

In this paper, we propose an ecosystem that incentivizes authors and reproducers to contribute to reproducible research. After discussing our own experiences in reproducible

research (§ 2), we introduce various building blocks that can be adapted to venue-specific profiles (§ 3). A key element is the Reproducibility Challenge. We compare our proposal with related approaches outside computer networking (§ 4).

2 AUTHOR EXPERIENCES

We share our experiences from both authoring reproducible research and reproducing existing publications.

2.1 Authoring Reproducible Research

We refined our approach when submitting reproducible papers over time. Originally, we created a reproduction bundle, consisting of code, data, and other relevant artifacts, along with the initial submission to a venue. Where submission tools did not accept artifacts, we uploaded data to a private (*i.e.,* only mentioned in the submission) URL. Uploading artifacts to a public repository is often undesirable at this point of time: First, results are usually confidential until publication. Second, source code often needs a certain amount of cleanup, structuring, and documentation before publication, which may not be ready yet at the initial submission. However, the preparation of the reproduction bundle along with the initial submission turned out as too early. Reviews hardly ever discussed the artifacts, probably because most venues do neither allow for extra time during reviewing phase nor make reproducibility an explicit part of review criteria.

When artifacts are not explicitly considered during paper reviewing, we now declare our reproducibility intent in the initial submission, but only bundle and publish artifacts upon acceptance. This also reduces rework, as modifications to final submissions may also affect artifacts.

2.2 Reproducing Publications

We faced the following problems when reproducing papers:

Author Unavailability. Authors were frequently unresponsive to email inquiries. This is not surprising as they may have new positions and responsibilities, may have a busy schedule, or may have lost access to raw data themselves. Helping others in reproducing their original work may require hours or days of additional effort with little return. We conclude that reproducibility should be possible without author cooperation. However, this ideal situation is rather unlikely, in particular in case of the first reproduction. Thus, we also conclude that we need incentives for authors to interact with reproducers.

Artifact Unavailability. In papers that promise reproducibility, links to artifacts may have become unavailable, especially those stored on personal websites. We conclude that artifacts must be accessible long-term, if possible in the same digital library as the publication.

Lack of Detail. Workshop, conference, and journal publications are not technical reports but concise syntheses of scientific insights. Hence, available space is dedicated to analysis, often falling below the level of detail to accurately reproduce the work. We conclude that precise technical descriptions should be part of the submitted artifacts.

Unclear Terminology and Expectations. Some publications claim reproducibility, but only allow partial reproduction, *e.g.,* of the measurement, analysis, or visualization processes. For a comprehensive reproduction, all aspects should be reproducible. We conclude that terminology and expectations should be clearly articulated in the call for papers.

3 PROPOSED BUILDING BLOCKS FOR A REPRODUCIBILITY ECOSYSTEM

In this section, we outline building blocks to establish an ecosystem of reproducible research. We argue that some building blocks require initial experience, hence a venue might move along several maturity stages or profiles.

SIGCOMM Reproducibility Challenge. To foster reproducibility in practice, we introduce the Reproducibility Challenge. The Reproducibility Challenge is a forum to reproduce published papers, potentially interact with authors of the original paper, and present the results. We suggest to co-locate this event with a flagship conference such as SIGCOMM. This has several advantages. (*i*) Reproducers gain visibility, which is an important incentive. (*ii*) Authors of the original paper likely attend the conference for other reasons, which reduces logistics and gives reproducers the chance to interact face-to-face. (*iii*) Discussing the outcome of the reproduction on a well-established conference will also give incentives to original authors to improve the reproducibility of their results.

This annual workshop at the SIGCOMM conference would accept reproductions of past publications, and also provide room for practical interactions to close gaps. To finish reproduction during the meeting, it could be structured as a two-day event, one at the beginning and one at the end of SIGCOMM. Participation should be free of charge (*e.g.,* sponsored by SIGCOMM) to encourage attendance. We suggest that written reports, which can be cited properly, are published as a result of the reproduction. This reproducibility challenge forms the most basic building block. It is easy to establish (co-located workshops are common at SIGCOMM), promotes the importance of reproducibility, and creates basic incentives (*e.g.,* visibility) for authors and reproducers.

Author Incentives. Having gained experiences on reproducibility, venues can provide additional incentives to authors for improving reproducibility. Starting with the call for papers, each venue should set its own expectations for reproducibility. While making reproducibility a requirement for submission may be too restrictive, a venue should create positive incentives for authors to provide reproducible papers. This might include a higher chance of acceptance (*e.g.,* by using reproducibility as a tie break), positive comments at the conference or editorial, or a best reproducibility award for past publications. The latter also incentivizes authors to readily answer reproducers' questions. However, when awarding reproducible papers, reproducibility must be reviewed timely.

Reproducibility Review. For venues that provide explicit incentives for reproducibility, we recommend a two-stage review process where submissions are first evaluated for their technical merit, and, upon acceptance, verified that the reproducibility claims hold. For this, we recommend establishing a SIGCOMM *Reproducibility Review Committee (RRC)* similar

Table 1: Recommended Use of Building Blocks along Reproducibility and Research Maturity.

Building Block	Initial	Evolved	Mature
Reproducibility Challenge	✓	✓	✓
Author Incentives	✘	✓	✓
Reproducibility Review	✘	✓	✓
Metrics & Badging	✘	✘	✓
Journal Fast-Track	✘	✘	✓

to an Artifact Evaluation Committee (AEC) [1], which could be a centrally organized pool of proficient graduate students, who gain recognition in the community from their participation. RRCs should be chaired by well-respected senior leaders in the community. The RRC can provide reproducibility reviewers to individual venues. After acceptance notification, a *reproducibility shepherd* is assigned to each paper, advising the authors towards a well reproducible camera-ready submission. This includes selection of a well-suited archival page, data formats, and basic checks whether the supplied artifacts are available at camera-ready time and plausibly appear to cover the claims made in the publication. However, as the short time between notification and camera-ready submission does not allow for in-depth reproduction of work, the expected result of this reproducibility review is a *ready for reproduction* badge, indicating that the reproducibility shepherd considers the provided artifacts suited to foster reproduction. We argue that detailed methodology should also be provided with the artifacts, not as yet another separate report. A good quality test is whether the provided artifacts would allow to confidently falsify the claims of a paper.

Metrics & Badging. After gaining some experiences with reviewing reproducibility, venues may award badges [3] to publications that meet certain criteria of reproducibility. Awarding badges requires metrics. For *replicability*, the metrics should be two-fold. First, the fraction of factual claims and actually provided artifacts should be calculated. Second, each of those data and code artifacts should be assessed with respect to completeness, *i.e.,* verification if the full process of measurement, pre-processing, analysis, and visualization is fully or partially covered. For *reproducibility*, comparison of the publication and independently gained results is necessary.

Journal Fast Tracking. Once the different levels of reproducibility are well-understood, they may inform journal fast-tracking. Also, special journals could require proven reproducibility as a unique feature.

We summarize the proposed building blocks and their applicability in Table 1. It is worth noting that the implementation of each building block depends on the type of venue. A venue focusing on novel ideas will likely want to avoid the formalizing and time-consuming building blocks, which, for example, journals might use. As a stage even beyond *mature*, *rigorous* venues might choose the desirable stance to only accept proven reproducible research for publication.

Table 2: Building Blocks Leveraged by Existing Approaches to Reproducibility.

Building Block	CCR	AEC	HSCC
Reproducibility Challenge	✘	✘	✘
Author Incentives	✓	n/a[1]	✓
Reproducibility Review	✓	✓	✓
Metrics & Badging	✓	✘	✓
Journal Fast-Track	n/a	n/a[1]	✘

1: Details depend on specific venue, not central AEC.

4 EXISTING APPROACHES

In computer network research, there is mainly no formal setting for reproduction of published papers. A notable exception is SIGCOMM CCR, which allows for more pages if a paper is submitted with artifacts. This journal follows a two-step review process, first reviewing technical content and then quality of reproducibility descriptions and artifacts. In parallel to our work, the Yearly Networking Contest [6] is proposed. The main difference is that the Reproducibility Challenge targets also on reproduction that requires more than a few weeks. Also in parallel, Bajpai *et al.* [5] discuss, among other issues, a lack of reproducibility incentives, and highlighting of reproducible papers as one solution approach. A Stanford course [9] integrates reproduction into education. An artifact-focused approach has been implemented by members of the programming language and software engineering community. Between notification and camera-ready deadline, a rolling *Artifact Evaluation Committee (AEC)* [1] basically validates whether the paper complies with the reproducibility claims, based on the provided artifacts. In case of successful review, a badge is awarded and can be included in the paper. AEC reviewers are mainly graduate students. Another artifact-focused approach is implemented at ACM HSCC Conference [2], which offers an opt-in service to assess and potentially award reproducibility of technical papers *after* acceptance. Assessment includes coverage, instructions, and quality of artifacts.

Table 2 maps existing approaches to our building blocks (§ 3).

5 CONCLUSION

We argued that the low number of reproducible papers is mainly due to the lack of incentives for authors and reproducers. We presented building blocks for an ecosystem that provides positive incentives to leverage reproducible research. A key element is the *SIGCOMM Reproducibility Challenge*, an event co-located with annual SIGCOMM conference, where authors and reproducers can meet to reproduce research and discuss results in a visible setting.

Acknowledgments. This work has been supported by the German Federal Ministry of Education and Research project X-Check, under grants 16KIS0528K, 16KIS0529, and 16KIS0530.

REFERENCES

[1] http://www.artifact-eval.org/, 2/2017.
[2] https://www.cs.ox.ac.uk/conferences/hscc2016/re.html, 3/2017.
[3] ACM. Result and Artifact Review and Badging. http://acm.org/publications/policies/artifact-review-badging, 1/2017.
[4] ACM. Workshop on Models, Methods and Tools for Reproducible Network Research. http://conferences.sigcomm.org/sigcomm/2003/workshop/mometools/, 2/2017.
[5] V. Bajpai, M. Kühlewind, J. Ott, J. Schönwälder, A. Sperotto, and B. Trammel. Challenges with Reproducibility. In *Proc. of ACM SIGCOMM Reproducibility Workshop*. ACM, 2017.
[6] M. Canini and J. Crowcroft. Learning Reproducibility with a Yearly Networking Contest. In *Proc. of ACM SIGCOMM Reproducibility Workshop*. ACM, 2017.
[7] M. P. Grosvenor, M. Schwarzkopf, I. Gog, R. N. M. Watson, A. W. Moore, S. Hand, and J. Crowcroft. Queues Don't Matter When You Can JUMP Them! In *Proc. of USENIX NSDI*, 2015.
[8] M. Wachs, Q. Scheitle, and G. Carle. Push Away Your Privacy: Precise User Tracking Based on TLS Client Certificate Authentication. In *Proc. of Network Traffic Measurement and Analysis Conference (TMA)*, 2017.
[9] L. Yan and N. McKeown. Learning Networking by Reproducing Research Results. *ACM SIGCOMM Computer Communication Review*, 2017.