Learning Reproducibility with a Yearly Networking Contest

Marco Canini
KAUST

Jon Crowcroft
University of Cambridge

ABSTRACT

Better reproducibility of networking research results is currently a major goal that the academic community is striving towards. This position paper makes the case that improving the extent and pervasiveness of reproducible research can be greatly fostered by organizing a yearly international contest. We argue that holding a contest undertaken by a plurality of students will have benefits that are two-fold. First, it will promote hands-on learning of skills that are helpful in producing artifacts at the replicable-research level. Second, it will advance the best practices regarding environments, testbeds, and tools that will aid the tasks of reproducibility evaluation committees by and large.

CCS CONCEPTS

• **Social and professional topics** → **Computing education**; • **Networks** → *Network performance evaluation*;

KEYWORDS

Reproducible research, Gamification, Design contest

ACM Reference format:
Marco Canini and Jon Crowcroft. 2017. Learning Reproducibility with a Yearly Networking Contest. In *Proceedings of Reproducibility '17, Los Angeles, CA, USA, August 25, 2017,* 5 pages.
https://doi.org/10.1145/3097766.3097769

1 INTRODUCTION

Concerns over irreproducibility of scientific results in the scholarly literature are raising discussions in several academic communities over methods and recommendations to incentivize and enhance reproducibility of research [14, 17, 21, 32]. To this end, several computer science communities have established a reproducibility review of papers. For example, the database community has had a repeatability committee at SIGMOD since 2008 and at VLDB since 2012 [21, 26]. The programming language community has established artifact evaluation committees since 2011 [25], which have been used in about two dozen conferences. Recently, the HPC community introduced a reproducibility initiative at SC 2016. While repeatability of experiments is not mandated, authors of accepted papers are encouraged to submit their codes and data to a committee, which typically consists of senior graduate students and postdoctoral researchers. Papers whose experiments can be repeated receive a badge or label to be shown as a distinguished recognition.

Doing reproducible research is without a doubt important for the integrity and betterment of scientific progress. Moreover, studies that reproduce prior work increase confidence in those results and expand on them. For example, a study by Clark et al. [16] confirmed the Xen hypervisor performed as expected but also presented new results on a less powerful PC, discovering a relatively comparable overhead. Another study by Howard et al. [24] assured that the original Raft paper was sufficient to allow independent re-implementation but also recommend several optimizations to the protocol.

Within the networking community, there is anecdotal evidence that reproducibility can have numerous benefits for the community (e.g., learning networking by reproducing research results [33]) but our current practices are still lacking and require improvement, in particular how to encourage broader participation as well as what tools and resources to use, and how to create, package and share reproducible experiments. In fact, although reproducibility would ideally be close to zero effort, it is actually hard work for both the authors of the original research – who have to produce artifacts solid enough to be shared – and the persons reproducing results – who need to spend time getting up to speed and dealing with missing details or documentation.

We propose to improve the extent and pervasiveness of reproducibility of networking research by organizing a yearly SIGCOMM contest, that is, an event of international scale, open to all students, and with the branding and prestige for which participants would receive external recognition. Our rationale is that a contest fosters a spirit of competition, which can play as an extra incentive to participation. This contest aims to enable students to learn how to perform reproducible networking research. To this end, the contest will require participants to learn and engage in reproducible experiments, improving their knowledge and skills. We believe students that are still at an early stage in their research careers, possibly well before they have any published work, may not yet have an appreciation towards reproducible research and would benefit from one such educative process. By gaining (hopefully enjoyable) experience in the contest, experiencing the process and potential benefits of reproducible experiments, and perhaps even winning some sort of prize or recognition, students might develop an appreciation for reproducible research and in the future, be more inclined to improve reproducibility for their own research. In contrast, senior students are more likely to be entrenched in their own (reproducible) research and with limited time to spare. The community may be better served if these students take part in reproducibility evaluation committees.

Our proposal does not substitute nor conflict but rather expands other processes like reproducibility evaluation committees and

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

ACM ISBN 978-1-4503-5060-0/17/08…$15.00
https://doi.org/10.1145/3097766.3097769

special-issue journals [20], as well as courses that make reproducing results an integral part of the coursework [33]. It is not our goal to establish a "best of reproductions" competition nor to add another checkbox for reproducible research. However, targeting students might improve the overall impact because (*i*) it can influence their research process before it has matured and (*ii*) they have their whole career ahead of them. A measure of success for us will be to observe an increase of reproducible research authored by students that participated in the contest, whether prize winners or not. We think that just having a reproducibility badge does not mean that more people know how to do reproducible research (even if they want to) whereas our proposal creates an environment where more people should end up having the right skillset and mindset. Thus, we expect that our success will propel the success of other processes, and be a cog in the wheel for making research reproducibility more pervasive.

The underscoring goal for many contests presently organized in several computer science communities is to advance the state of the art — e.g., tackling a big challenge problem like creating autonomous cars as in the DARPA Grand Challenge [4]. Such a goal is partly at odds with improving research reproducibility; moreover, there does not seem to be an agreement in the networking community about its grand challenges [22, 28]. We review below existing contests in networking and find that their formats and/or goals also do not integrate well with enabling research reproducibility.

One possibility could be to organize a contest where participants have to reproduce or replicate experiments from published papers. This format has several possible benefits, including that research results will be verified independently and participants who are actively trying to replicate experiments will likely face hard, technical challenges and, as a result, become aware of and be able to make recommendations on how to better package experiments with code, data, and documentation. Participants may be incentivized with awards and prizes; additionally, there are incentives for the original authors who would like to see their results reproduced to have higher impact and visibility for their works. This approach would meet our goal. However, contests have winners, and it remains unclear what criteria to use to appropriately and objectively rank participants, especially if the percentage of successful reproductions is high.

Another possibility is to organize a contest in a format more similar to a programming or design contest where problems are well defined and have quantitative goals (e.g., develop the "best" congestion-control protocol in certain settings [31]). Having measurable and unambiguous performance indicators is beneficial in that it enables to direct the search for a solution and to score participants' submissions objectively. However, reproducing research is generally not an integral aspect of this type of contest.

Given our current goal, we propose to design the contest as a middle ground between these two possibilities. The objective should not be on replicating an experiment from published work per se. That said, a first phase of the contest may consist of creating a baseline solution by building upon an existing approach and reproducing a certain experiment from a paper. In a second phase, the focus may shift to finding a solution that performs better than the baseline. In a sense, the contest could be designed mirroring a pedagogical view that promotes creating clean and well-structured solutions that need to be packaged and shared with the contest evaluators, creating a disincentive for the "hack it together" mentality. We believe this will make reproducibility an integral part of the contest. The contest will take reproducibility of research a step further by not only reproducing existing research but also advancing upon the state-of-the-art and "learning reproducibility" as a skill as a by-product of participating in the contest.

With the above in mind, the contest should also be designed with consideration for how much time is allowed. The format of an hackathon is not appropriate due to its short duration. Also, we believe the contest should be organized as a team challenge to allow a small group of students to work together towards a substantial problem. It should be repeated yearly, and there should be a rotation of the organizers (e.g., a research group) to ensure fresh ideas and energy is injected each year as well as sharing the load undertaken by the organizers.

In the following, after a brief summary of existing contests in networking, we elaborate on how to connect a contest with learning research reproducibility. We review examples of courses at notorious universities where traits of our proposal have been applied and then suggest a possible framework for our contest.

2 CONTESTS IN NETWORKING

In the networking industry, the Cisco Networking Academy Contest [2] and Jubilee 15th Challenge24 Contest Networking [5] focus on practical hands-on experience and skills with current networking technology rather than research-oriented problems. The Wireless Battle of the Mesh [11] is an event aimed at networking enthusiasts and community networking activists to come together to test the performance of different routing protocols for ad-hoc networks. With the advent of SDN, the contests that have been organized have primarily focused on problems relevant to the industry. The Open Networking Foundation held the OpenFlow Driver Competition [7], which narrowly focused on OpenFlow driver development. The ONUG Grand Challenge Hackathon [6], while focusing on the broader scope of network automation, was organized in the format of a hackathon located at the ONUG meeting.

The closest approximation to an academic-oriented international contest is perhaps the 2015 AT&T's SDN Network Design Challenge [1], in which participants had to improve network routing for a realistic, carrier-grade network. While the contest emphasized openness and contributions to open source, only U.S. citizens were eligible to participate. Interestingly, these examples, as well as others from academia that we relate to in more details below, indicate that networking as a problem area admits many possibilities to create sufficiently well-defined problems that anyone with the right background can attempt as part of a contest.

However, as far as we know there has never been an equivalent of a "SIGCOMM contest," that participating students would receive external recognition for. Particularly, there has been no contest for the purpose of encouraging reproducibility. A concurrent paper to ours by Scheitle et al. [30] proposes a Reproducibility Challenge to be co-located at SIGCOMM to incentivize original authors and reproducers; however, this event is not in the form of a contest.

We suggest that our proposal and theirs should be combined to maximize impact in the community.

3 HOW DOES A CONTEST AID WITH RESEARCH REPRODUCIBILITY?

We argue that a contest properly designed around the goal of promoting better reproducibility of research will have the following benefits:

First, participants will learn something of the skills, methods and technologies that are broadly useful in creating reproducible research. Participants will need to reproduce an experiment and also to make their own work reproducible, which will hopefully result in them gaining a greater understanding of what is required to reproduce an experiment and make it reproducible. For example, the contest may require students not only to adopt a certain network emulator or simulator, set up an experimental environment that includes realistic traffic generation, work with some building blocks like Linux containers, version control, etc. but also to create and package their solutions in a way that these can be shared with the contest evaluators and can be executed on their test harness. While a contest is by no means the sole way to learn about reproducibility, the excitement and challenges brought by a contest will further stimulate students, as they focus on tackling a well-defined problem, to learn and develop experience with methods and tools that the community expects to meet certain "golden standards" of conducting reproducible research [13]. It is then expected that participants will develop an appreciation towards reproducible research.

Second, organizers will gain experience with creating test harnesses, testbeds, datasets, and platforms for evaluating the participants' submissions. The evaluation of submissions in itself is a challenging problem. The experiences and lessons learned are valuable to the networking community in that they can help to implement and evaluate models and systems for reproducing and validating experiments – i.e., advance the "best practices" for research reproducibility evaluations. This will hopefully result in tangible benefits for research reproducibility reviewing committees and support the sociotechnical system, including archives, testbeds, budget, and staff, that could facilitate ongoing replication and validation of networking research results. For instance, certain research unfortunately cannot be reproduced because it requires a particular hardware or testbed that might not be available or software and data that cannot be shared. However, as part of running the contest, organizers might develop infrastructure that is reusable and applicable to a broader scope than the specific requirements of the contest. In certain cases, this might offer a solution to reproducing some of the hard-to-reproduce research.

Is there any example that supports the above argumentation? Although several researchers are building methodologies and tools for designing and deploying repeatable experiments on network testbeds (e.g., see [19, 29]), experiences with reproducibility evaluations in the networking community are still limited. For instance, it is only since January 2017 that the SIGCOMM Comput. Commun. Rev. (CCR) has adopted a process to publish long papers whose results can be repeated [15]. However, it is interesting to see that there exists anecdotal evidence from settings where reproducing research results has led to both the learning of technical coding and experimental skills as well as the creation of tools in support to reproducing results. We now review some of these examples.

Stanford CS 244 example. As a point in case, consider the example of Stanford CS 244, which is a course on Advanced Topics in Networking. Since 2012, students taking this course perform a group assignment in teams of two students with the goal of replicating research results published by other researchers from a paper of their choice. Quoting from [9], it is especially interesting to note that students are asked to:

"Replicate an existing result (good), show a limitation of the chosen platform that prevents you from replicating the result (negative results are equally good), or challenge the result in the paper with data (better), or produce a new result (even better)".

"Implement experiments in a way that is easily replicated by others– ideally, another researcher can install it via a single script command, run it with a single command, and generate output graphics with a single command".

The final deliverable requires that the students write a blog post on a dedicated website [8] and make their code available.

Considering the past five editions of the course since 2012, the website shows that over 100 teams have posted their projects. Interestingly, it shows a combination of both successfully reproduced experiments (with a high frequency) and failed ones (with a low frequency). Each blog post describes in details the experience of reproducing a certain result, including technical challenges, platforms, software dependencies, execution environments, and instructions for reproducing the students' experiments. This body of documentation is a testament to the large work that happened behind the scenes and what was learned as a result of going through the process (and trouble) of reproducing the chosen paper by each team.

For the interested reader, a recent CCR editorial [33] describes in details the experiences of Stanford CS 244 students regarding the interplay of learning networking by reproducing research results. Interestingly, the authors observe that *"reproducing research can simultaneously be a tool for education and a means for students to contribute to the networking community"* [33].

ITN METRICS example. Another example that the authors are aware of is the "bootcamp" of the Marie-Curie ITN METRICS project [12] where a sizable group of students got acquainted with Internet measurements research by reproducing results from papers during a two-week long event at Université catholique de Louvain. The short period available meant that measurement papers were selected based not only on their importance but also primarily because they could be reproduced within a two-week period by using existing datasets. Without any doubt, the experience of the students has been relevant for them to learn "hands-on" and cope with the challenges of reproducing Internet measurements research while working cooperatively in small teams.

These examples, however relevant from the reproducibility perspective, did not take place in the context of a contest. A further example provides anecdotal evidence of the benefits of a contest towards reproducibility.

MIT 6.829 example. Students of MIT's graduate course 6.829 in computer networks participated in 2013 as part of the coursework in a contest to develop congestion-control protocols for cellular wireless networks. An article by Sivaraman et al. [31] documents

the experience of the course staff and the infrastructure that was created while running the protocol design contest as part of the course. Based on their experience, the authors draw the following among their conclusions about contests: *"if designed properly, such contests could benefit networking research by making new proposals more easily reproducible"* [31]. Moreover, they identify the following rationale as to why such a kind of contest aid with research reproducibility: *"turning a research problem into a well-specified contest forces the researcher to clearly articulate the testing conditions and ensure that her protocol works reproducibly under those conditions. This, in turn, makes the protocol accessible to a wider audience of other researchers"* [31]. The congestion protocol contest appears to have been a successful experience and his now being repeated yearly since 2015 as part of Stanford CS 344G [3] and has been recently repeated in MIT 6.829 in 2016.

Tooling examples. Finally, it is worth noting that efforts to engage students with reproducibility of networking results have influenced tools such as Mininet HiFi [23] and SELENA [27] that help to emulate network conditions and to be able to execute artifacts in controlled environments. The MIT 6.829 course staff developed a substantial infrastructure to enable student submissions to be portably and repeatably tested as well as to allow each student team to evaluate their protocols under unobserved conditions [31].

4 WHAT FORMAT FOR THE CONTEST?

The desired outcome of the contest is to instill in students an appreciation for reproducible research as well as a learning experience of what reproducibility entails. We now discuss a possible contest format to achieve this outcome, which is loosely based on the protocol design contest [31]. Admittedly, how to best structure such a contest remains an open question.

It seems natural to allow students to compete as part of small teams. The duration should be in the order of a few weeks to allow for students with sufficient background to jump into the problem and attempt it. We assume the contest will last for four weeks in practice; however, this can be modulated depending on the context. The topic of the contest (e.g., SDN, data center networks, wireless, inter-domain routing, etc.) as well as the paper that it is based upon, will be advertised in advance to allow students decide whether to enter it; however, the details will only be published starting with the beginning of the contest. Throughout the contest, participants might receive partial credits that are accumulated to determine the final rankings, but submissions will be primarily judged based on their overall performance on a final workload, which is not shared with participants to avoid "overfitted" solutions.

At Week 1, participants will receive a skeleton code and get acquainted with the required experimental environment through a small set of exercises. At Week 2, participants will need to reproduce an experiment (or a few ones) for the paper that the contest is based on and submit their solution for validation on a supplied workload. Besides familiarizing participants with the basics of how to reproduce an experiment, these two initial weeks should help students to understand certain aspects of the original work in sufficient detail that will be helpful for them to try improve on the work despite being exposed to it for only a few weeks overall. At Week 3, the contest enters a second phase wherein participants need to improve upon the existing, validated and reproduced solution. During this week, they will be able to submit their current solution for a certain number of attempts (possibly unlimited) and receive feedback through a scoreboard about how well their solution fairs against others. The scoreboard may reflect a coarse-grained performance metric or specific indicators that can be useful to improve a team's solution. At Week 4, participants will fine tune their solutions for a final round of evaluations. During this time there should be just a minimal feedback to aid the participants and let the best of breed solution standout.

Ensuring a smooth running of such a contest is not without challenge and likely will require careful planning and constant effort throughout the contest duration. Via rotation, a research group in a given year should have ownership of organizing the contest. The process may include a proposal phase if many groups offer to be the organizers. The organizing team would pick the paper that the contest is based on since it requires familiarity with the work and the ability to coordinate evaluation and validation of submissions. Though, with more maturity of the process, the selection may be influenced by recent papers that were awarded with a reproducibility badge, incentivizing participations of original authors to various reproducibility initiatives through the additional opportunity for their works to broaden the impact via the contest.

To further incentive partition, the contest may follow an approach similar to the yearly SIGMOD contest: the top three submissions will be invited to compete in a "bakeoff" to be held at SIGCOMM where the prizes are awarded; up to two students from each team will receive travel grants to attend the conference. This stage might be integrated as part of the SIGCOMM Reproducibility Challenge [30].

The winning team will be awarded a prize, which would be donated by companies. Competitions' prizes need to be sufficiently visible that participants can include them on their CV, so that they become a new metric for good work. Currently, external recognition is largely measured by published papers. However, there are tendencies in fast moving and exciting areas to place weight on originality at the expense of rigor or else due to the importance of the topic to industry, to work in large teams on the next problem, and emphasize thoroughness at the expense of reproducibility. What a prize does is to provide public recognition that reproducing results is rigorous science and engineering, which should stand at least equal to originality or impact/significance.

Moreover, the winning submission (or say the top three) may receive an invitation to publish an article on CCR about their reproduced results and improvements, along with their codes. Over time, a committee might consider these articles for a category of "best reproducible research" and select among them for an award just like the test of time award, or CACM research highlights. The original paper might be included in the award to also incentivize original authors to propose and organize a contest around their research.

5 WHERE DO WE GO FROM HERE?

This paper aims to start a discussion around the concept of a yearly international contest to be organized within the SIGCOMM community. It is hoped that this event will substantially aid with ongoing

efforts to promote reproducibility of research results in networking by offering a framework for hands-on experience with creating and sharing artifacts that solve well-defined networking problems evaluated through remote testbeds on challenging workloads and datasets, thus effectively requiring *replicability* by an independent team. We invite everyone interested to contribute ideas and offer assistance in building a critical mass to advance to the next stage where concrete actions can be planned.

In particular, we hope the workshop will be a good opportunity to discuss additional ideas regarding the format of contests and how to define tractable, interesting problems that feed into the overarching goal of boosting research reproducibility. Examples might be congestion-control over some modeled flow arrivals and network, or multi-node wireless transmission over some medium, or speed or memory required to make some kind of routing decision, or even an FPGA design to switch a certain number of packets/sec with a constrained amount of silicon. However, one challenge in identifying topics might be that the community is broad in terms of its interests and topics, as networking is clearly a rich problem domain but perhaps not yet a scholarly discipline in itself [28].

The exercise of developing contests may in itself have unexpected positive outcomes and impact. For instance, the organizing team may be creating interesting workloads and benchmarks for networking problems. A similar initiative, called VideoNet [10], has created a community of computer vision researchers that have put effort into creating benchmarks for video-related tasks. The networking community could attempt to have a similar community to foster the exchange of benchmarks and ideas, and the contributions of the contest organizers would be valuable. In addition, developing contests might be a relevant motive for the community to reflect on and identify networking problems and topics where consensus exists regarding their importance. It is our hope that reaching such a consensus may further aid the community towards developing a prescriptive network theory, which at the moment remain elusive [18].

Acknowledgments. We thank our shepherd, Bob Lantz and the reviewers for their feedback. We are thankful to Olivier Bonaventure, Luigi Iannone, David Keyes, Shriram Krishnamurthi, Jennifer Rexford, Robert Ricci, Damien Saucez, Matthias Waehlisch, and Keith Winstein for sharing useful references and giving us suggestions and support for this paper.

REFERENCES

[1] AT&T's SDN Network Design Challenge. http://about.att.com/innovationblog/051515acallouttoacad.
[2] Cisco Networking Academy Contest. http://www.academynetriders.com/.
[3] Congestion-Control Contest. https://web.stanford.edu/class/cs344g/contest.html.
[4] DARPA Grand Challenge. http://www.darpa.mil/grandchallenge/.
[5] Jubilee 15th Challenge24 Contest Networking. http://ch24.org/networking.html.
[6] ONUG Grand Challenge Hackathon. https://opennetworkingusergroup.com/onug-grand-challenge-hackathon/.
[7] OpenFlow Driver Competition. https://www.opennetworking.org/competition.
[8] Reproducing Network Research. https://reproducingnetworkresearch.wordpress.com/.
[9] Stanford CS 244 – Programming Assignment 3. http://web.stanford.edu/class/cs244/pa3.html.
[10] VideoNet – A benchmark initiative for all things video. http://videonet.team/.
[11] Wireless Battle of the Mesh. http://battlemesh.org/.
[12] METRICS Bootcamp. http://metrics-itn.eu/metrics-bootcamp/, 2014.
[13] New gold standard established for open and reproducible research. http://www.cam.ac.uk/research/news/new-gold-standard-established-for-open-and-reproducible-research, 2015.
[14] R. F. Boisvert. Incentivizing Reproducibility. *Commun. ACM*, 59(10), Oct. 2016.
[15] O. Bonaventure. The january 2017 issue. *SIGCOMM Comput. Commun. Rev.*, 47(1), Jan. 2017.
[16] B. Clark, T. Deshane, E. Dow, S. Evanchik, M. Finlayson, J. Herne, and J. N. Matthews. Xen and the Art of Repeated Research. In *ATEC*, 2004.
[17] C. Collberg and T. A. Proebsting. Repeatability in Computer Systems Research. *Commun. ACM*, 59(3), Mar. 2016.
[18] J. Crowcroft, M. Fidler, K. Nahrstedt, and R. Steinmetz. Future Internet (Dagstuhl Seminar 13131). *Dagstuhl Reports*, 3(3), 2013.
[19] S. Edwards, X. Liu, and N. Riga. Creating Repeatable Computer Science and Networking Experiments on Shared, Public Testbeds. *SIGOPS Oper. Syst. Rev.*, 49(1), Jan. 2015.
[20] E. Eide. Foreword: Special Issue on Repeatability and Sharing of Experimental Artifacts. *SIGOPS Oper. Syst. Rev.*, 49(1), Jan. 2015.
[21] J. Freire, P. Bonnet, and D. Shasha. Computational Reproducibility: State-of-the-art, Challenges, and Database Research Opportunities. In *SIGMOD*, 2012.
[22] P. B. Godfrey. What are Networking's Grand Challenges? http://conferences2.sigcomm.org/co-next/2015/#!/keynote, 2015. CoNEXT keynote.
[23] N. Handigol, B. Heller, V. Jeyakumar, B. Lantz, and N. McKeown. Reproducible Network Experiments Using Container-based Emulation. In *CoNEXT*, 2012.
[24] H. Howard, M. Schwarzkopf, A. Madhavapeddy, and J. Crowcroft. Raft Refloated: Do We Have Consensus? *SIGOPS Oper. Syst. Rev.*, 49(1), Jan. 2015.
[25] S. Krishnamurthi and J. Vitek. The Real Software Crisis: Repeatability As a Core Value. *Commun. ACM*, 58(3), Mar. 2015.
[26] I. Manolescu, L. Afanasiev, A. Arion, J. Dittrich, S. Manegold, N. Polyzotis, K. Schnaitter, P. Senellart, S. Zoupanos, and D. Shasha. The Repeatability Experiment of SIGMOD 2008. *SIGMOD Rec.*, 37(1), Mar. 2008.
[27] D. Pediaditakis, C. Rotsos, and A. W. Moore. Faithful Reproduction of Network Experiments. In *ANCS*, 2014.
[28] J. Rexford. The Networking Philosopher's Problem. *SIGCOMM Comput. Commun. Rev.*, 41(3), July 2011.
[29] R. Ricci, G. Wong, L. Stoller, K. Webb, J. Duerig, K. Downie, and M. Hibler. Apt: A Platform for Repeatable Research in Computer Science. *SIGOPS Oper. Syst. Rev.*, 49(1), Jan. 2015.
[30] Q. Scheitle, M. Wählisch, O. Gasser, T. C. Schmidt, and G. Carle. Towards an Ecosystem for Reproducible Research in Computer Networking. In *Reproducibility*, 2017.
[31] A. Sivaraman, K. Winstein, P. Varley, J. a. Batalha, A. Goyal, S. Das, J. Ma, and H. Balakrishnan. Protocol Design Contests. *SIGCOMM Comput. Commun. Rev.*, 44(3), July 2014.
[32] V. Stodden, M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, M. A. Heroux, J. P. Ioannidis, and M. Taufer. Enhancing reproducibility for computational methods. *Science*, 354(6317):1240–1241, 2016.
[33] L. Yan and N. McKeown. Learning Networking by Reproducing Research Results. *SIGCOMM Comput. Commun. Rev.*, 47(2), Apr. 2017.