Let's separate dissemination from selection

Stavros Tripakis
University of California, Berkeley
`stavros@eecs.berkeley.edu`

August 3, 2011

What

I suspect that scientific publications originated primarily as a means for disseminating research results. Today we live in a bigger and more complex, and perhaps also more number-dominated world, and publications serve also the goal of ranking researchers according to productivity, impact or other metrics [1, 2]. We also live in a highly-connected world, via the internet and the web. I believe that we can leverage this technology in ways that, as far as I know, we have not yet seriously pursued, in order to improve research dissemination. In the long run, these changes could also result in improving review, selection and scientific debate.

Dissemination, review and selection are distinct functions of the publication process. The goal of dissemination is to spread research ideas and results throughout the community. The goal of peer review is to check those ideas, in terms of correctness, originality, or other metrics, and provide feedback to the authors so that they can improve their work.[1] The goal of selection is to select those ideas and results that seem more worthwhile. Conferences and journals today perform all three functions. Dissemination and selection are usually not separated in the publication process: a paper that does not get selected by the program committee of a conference will not be included in the conference proceedings, therefore will not be disseminated (at least not through this channel).

I suspect that the reason for not separating selection and dissemination is historical and pragmatic: in the pre-internet days, paper proceedings or journals were the only means of publication. Only a limited number of articles can be included in a proceedings book or a journal issue, therefore a selection has to be made. Naturally, the articles that are not selected, are not published, and not disseminated either. Today, with electronic publications, this resource limitation no longer exists: the on-line (or even CD-ROM) proceedings or journal can include a virtually unbounded number of articles.

Of course, this does not mean that review and selection are no longer necessary. Peer review is a pillar of scholarly work. Selection is also necessary for many different reasons. One reason is that, in conferences, there is still the limitation of time: in a meeting that lasts three days, only a limited number of works can be presented. Therefore, some selection needs to be made.

---

[1]One can say that one of the roles of peer review is also to "protect from junk", i.e., filter-out those papers that are wrong, misleading, etc. This is particularly important in cases where such papers may mislead important decisions (e.g., think of the debate on climate change). I agree that protection from junk is important, but as with any protection mechanism, too much of it is not good.

How can dissemination and selection be separated? My proposal involves primarily conferences. It consists in giving authors the possibility to disseminate their work by providing links to this work through the conference web site, even if the work is not selected to appear in the proceedings or presented at the conference. Simply put, once review is complete, the conference site would have links to two lists of papers: *Selected papers* and *Submitted, not selected, but maybe still worth looking at, papers.* We will call the latter the SUB list. Naturally, the SUB list will not include a paper unless its authors want it to be included. Also, the SUB list need not include the actual papers (PDFs) but only links to those papers, or other ways to get access to the work.

Since first writing this, and the remarks below which have not been updated, I was happy to find that others are also proposing similar things. For instance, S. Keshav, Editor of ACM CCR, writes in [3]:

> "We can also remove the notion of a publication bar altogether. An online version of CCR, which will be coming some day, could easily accept all articles submitted to it. Editors and reviewers could rank papers and do public reviews and readers can judge whether or not to read a paper. This is already common practice in physics, using the Arxiv system."

Also, some venues are taking steps that can be seen as going in a similar direction. For instance: The *Fun Ideas and Thoughts (FIT)* workshop at the 2009 Conference on Programming Language Design and Implementation (PLDI) (http://people.csail.mit.edu/rabbah/conferences/09/pldi/fit/) makes available online a set of "Presented ideas" (paper, slides and blog) and a set of "Other submitted ideas" (paper and blog). The *2nd USENIX Workshop on Hot Topics in Parallelism (HotPar '10)* (http://www.usenix.org/events/hotpar10/) has selected 35 papers: all 35 papers will appear in the proceedings and all 35 will be accompanied by posters presented at the conference, but only 16 of those papers will have full presentations. The *Design, Automation and Test in Europe (DATE)* conference typically has sessions with long (30 min) and short (15 min) presentations, but both have papers with the same number of pages (6) in the proceedings. The *ACM/IEEE First International Conference on Cyber-Physical Systems (ICCPS 2010)* has, in addition to regular sessions, a poster session. Posters are accompanied with 1-page abstracts included in the proceedings.

Yes, but ...

Who will opt to disseminate a rejected paper? This is an interesting question, to be answered ultimately through experimentation: trying the method in some conferences for some time and see if it actually makes any difference.

Can't we do this already, via personal web sites and repositories such as arxiv.org or citeseer? Most authors have a web site, so they can put their paper on-line if they wish. Or they can archive it into on-line repositories such as arxiv.org, hal.inria.fr, etc. Other sites, such as DBLP or researchr.org generate such repositories automatically or semi-automatically, from published proceedings. These facilities do not serve the same purpose as making a paper accessible through the web site of a conference. Personal web sites and on-line repositories provide primarily a *search* service: if I know what I'm looking for, e.g., a paper by author X, or a paper with title Y, or a paper on topic Z, I can visit the web site of X, or google Y or Z, or search for them in a repository. Looking at the papers that have been selected at a conference, or submitted but not selected, provides also

a *browse* service: I do not know what I am looking for, I am just browsing the titles in case I find something that seems interesting. This is a fundamental difference, which makes existing services inadequate.

Can't we simply email our paper to our colleagues or announce it to some specialized email list? Again, these methods achieve different goals. Email is a direct method and can only be done in a limited way so that it does not become spam. Consulting a SUB list, on the other hand, is a browsing action initiated by the reader, not the writer.

Aren't 2nd-tier conferences, workshops, work-in-progress sessions, etc., serving the same purpose? Not really. First, there is delay in resubmitting a paper. 6 months to a year can be significant in a fast-moving field. Why not disseminate more quickly? Second, reject does not always imply non-mature works, works-in-progress, or works unworthy of dissemination. Finally, appearing in a SUB list does not preclude resubmission.

What you propose will create too much spam. If everyone has the right to disseminate anything by anchoring it to the web site of a prestigious conference, this might create incentive to receive a huge number of submissions, of low quality, or completely unrelated to the conference, etc. How to prevent spam? First of all, the SUB list need not include all papers submitted to the conference (although the idea is to include most of them, see below). If a paper is clearly out of scope, seriously erroneous, of extremely low quality, or inappropriate in any other respect, the program committee chairs or members can decide not to include this paper in SUB. Such serious problems are not the case for most papers that get rejected from a conference: the usual rejection message reads like *"We regret to inform you ... This year we received a record number of very high quality submissions ..."*. Also note that SUB lists are optional to consult: if necessary, readers could avoid dissemination spamming by simply choosing to ignore SUB lists.

Second, if it is necessary to limit the number of submissions themselves, some type of controlled registration mechanism may have to be applied. Many on-line submission systems today require some type of registration, although it is typically uncontrolled (i.e., anyone can register). Controlled registration could be implemented, for example, by some type of invitation mechanism, where in order to register one has to be invited by someone who is already registered. In that case, professors could invite their students, colleagues could invite other colleagues, and so on. This process, if necessary, should not entail too much overhead, since there is only a limited number of conferences in each field, and registration needs to happen only once for each. Therefore, submission spamming, if it does happen, can be also avoided.

What you propose creates extra layers of reviewing and extra work for program committees. Who will decide what goes into SUB lists and what doesn't? PC chairs and members. These people have enough work already selecting the papers to appear at the conference, so why add extra work? I think this extra work should be negligible. Ideally most/all papers should be included in a SUB list (assuming their authors want this). To be unacceptable for the SUB list, a paper must have some very serious flaws, for instance: be completely out of scope; be seriously incorrect; reinvent the wheel[2]; be unreadable. It is important that *all* reviewers think the paper has indeed these serious flaws: if there are doubts, then the paper should be given a chance, as this is the whole point behind SUB lists. If PC chairs or reviewers are confident enough to mark a paper "not even

---

[2]To be sure, the set of papers that "reinvent the wheel" is not straightforward to capture. It should include papers that rehash what every knowledgeable researcher knows already, but this shifts the difficulty in interpreting "knowledgeable". Unfortunately, papers are often rejected for not proposing anything "new", although some of these papers contain interesting reformulations that shed new light to a topic.

worthy to be in the SUB list" then they should do so. Otherwise, they should give the paper the benefit of doubt.

Regarding implementation of the mechanism, one may imagine an option "Unacceptable for SUB list" included in the review form, much in the same way as an option like "Consider for best paper award" appears in some review forms today. By default the option is unset. If all reviewers set the option, the paper can be excluded from the SUB list after discussion with the chairs.

What about original work requirements and copyrights? If we disseminate our work via SUB lists, won't this prevent the work from being published later on, since many conferences and publishers require that papers be original and subject to copyright? This potential problem can be avoided in the same way that it is avoided today in the case of pre-prints and technical reports. Often, researchers will write such a report and disseminate it in order to get feedback before actually submitting their work to a conference (in fact, in some disciplines, this seems to be a requirement in order to submit). Such pre-prints are earlier versions of the work that do not violate originality requirements. Ultimately, the role of copyrights in modern scientific publishing needs to be re-examined [4, 5, 6].

What about double-blind requirements? Many conferences require double-blind reviews, i.e., that author identities not be disclosed in the submitted paper. Does the appearence of a previous version of the work in a SUB list violate this requirement? This is unclear, as is unclear to what lengths the authors need to go in order to try to "hide" their identities in such systems. For example, does publishing a technical report violate this requirement? Does putting a pre-print on the institute's web site violate it? Does reporting on an implementation within framework X that everyone knows is being primarily developed by group Y violate it? The double-blind reviewing system is imperfect and at times it appears absurd.

Can we cite work that is listed on SUB lists? Yes, in the same way that technical reports, unpublished documents, or "personal communications" can be cited.

Why

Modern technology makes separating selection from dissemination possible, at almost no cost. But why would we want to pursue this option in the first place?

First, because doing so increases the number of ideas that circulate in a community, and makes such circulation much more timely. This is especially beneficial in a fast-moving field like CS.

Second, because it gives works a chance to be disseminated despite not being selected, thus making them more immune to some of the perverse effects of today's review, selection and evaluation processes [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. In the long run, separating dissemination from selection could shift part of the emphasis from the latter to the former, and could result in a cultural change with positive ramifications. For instance, works in progress, papers that convey new but incomplete ideas, or even failed attempts, all examples of elements that are useful to scientific progress but typically have difficulties finding a place in selective conferences, could be easily included in SUB lists.

Third, because it opens the way to other creative ways of leveraging internet and the web in order to improve the scientific research process. One can imagine, for instance, allowing readers to post comments about a paper on-line, similar to an on-line review (e.g., as in ACM's digital library, www.pubzone.org, etc.). These comments could be anonymous or eponymous, public or private. Authors can reply to the comments, creating a public or private debate. One could also imagine

using the same framework to establish links to related papers, tools, experimental data, and so on.

Such features can be technology enablers that could turn a conference web site into something more: an on-line forum, an on-line scientific community, as envisioned also in [8]. This could change in a profound way how research work is disseminated and debated in our field.

Note

As I get feedback from people or read others' ideas on this topic, I'm updating this document. This unfortunately results in some parts of the document being written earlier or later than other parts.

Acknowledgments

Thanks to all who take time to read this and special thanks to those who provide feedback.

References

[1] J. Ren and R. Taylor. Automatic and versatile publications ranking for research institutions and scholars. *Comm. ACM*, 50(6):81–85, 2007.

[2] B. Meyer, C. Choppy, J. Staunstrup, and J. van Leeuwen. Research evaluation for computer science. *Comm. ACM*, 52(4):31–34, 2009.

[3] S. Keshav. Editor's message. *ACM SIGCOMM Computer Communication Review*, 41(3), July 2011.

[4] R. Di Cosmo. Scientific Publications: The Role of Public Administrations in the ICT Era. *UPGRADE, The European Journal for the Informatics Professional*, VII(3), June 2006. http://www.upgrade-cepis.org/issues/2006/3/up7-3DiCosmo.pdf.

[5] M.Y. Vardi. Open, closed, or clopen access? *Comm. ACM*, 52(7):5–5, 2009.

[6] M. Beaudouin-Lafon. Open access to scientific publications. *Comm. ACM*, 53(2):32–34, 2010.

[7] D.L. Parnas. Stop the numbers game. *Comm. ACM*, 50(11):19–21, 2007.

[8] J. Crowcroft, S. Keshav, and N. McKeown. Scaling the academic publication process to internet scale. *Comm. ACM*, 52(1):27–30, 2009.

[9] K. Birman and F.B. Schneider. Viewpoint program committee overload in systems. *Comm. ACM*, 52(5):34–37, 2009.

[10] M.Y. Vardi. Conferences vs. journals in computing research. *Comm. ACM*, 52(5):5–5, 2009.

[11] L. Fortnow. Time for computer science to grow up. *Comm. ACM*, 52(8):33–35, 2009.

[12] J.M. Wing. Breaking the cycle. http://cacm.acm.org/blogs/blog-cacm/38402-breaking-the-cycle/fulltext, August 2009.

[13] J.M. Wing and M. Guzdial. CS woes: deadline-driven research, academic inequality. *Comm. ACM*, 52(12):8–9, 2009.

[14] International Symposium on Peer Reviewing (ISPR'09). http://www.ictconfer.org/ispr, 2009.

[15] S. Peyton-Jones. Two phase reviewing for POPL; a response. http://lists.seas.upenn.edu/pipermail/types-announce/2010/001740.html, January 2010.

[16] D.S. Wallach. Rebooting the CS Publication Process. http://www.cs.rice.edu/~dwallach/pub/reboot-2010-06-14.pdf, 2010.

[17] J.M. Wing and E.H. Chi. Reviewing peer review. Communications of the ACM, vol. 54, no. 7, pg. 10-11, 2011.