

Overcoming Challenges of Maturity

Kenneth P. Birman

Dept. of Computer Science, Cornell University

Abstract

In 2008, the systems and networks research communities find themselves victims of their own successes. This white paper reviews some of the evidence that the two areas are under enormous stress, and suggests that the situation is only going to get worse. Yet there are a number of simple steps we can take to reduce stresses if we start by asking ourselves what motivates the various parties involved. By fine-tuning the system, we can make life easier for ourselves, improve the level of satisfaction for typical researchers, open the door to innovative younger people who are thinking “out of the box”, and reduce the risk of abuses stemming from frustration.

1. Creaking at the Seams

The research community in systems and networks is showing increasing evidence of a dangerous form of stress. A vast boom in information technologies has already transformed the world, and yet is seemingly still in its early days. As this boom continues, participants will want to publish on their ideas. Visionary concepts such as NSF’s GENI initiative make it plausible that we’ll soon be reinventing the Internet, securing critical infrastructure and building applications capable of scaling to previously inconceivable degrees. One can easily anticipate that these developments will inspire all sorts of novel approaches to genuinely important questions. A tsunami of papers will surely result – overwhelming a conference system that is already creaking at the seams.

Meanwhile, our community holds just a handful of top-ranked conferences annually, and those have finite capacity for papers: finite in terms not just of the number that can be published annually, but also in terms of our ability to review submissions. Adding conferences can’t be the solution: by definition, they won’t be considered to be first-rate, and submissions still need to be reviewed. Yet if we *don’t* add more conferences, how in the world will all of these great ideas become known? The most frightening aspect, for those of us who maintain a high quality standard, is the prospect of needing to weed out an ever increasing number of papers that may be good, ok, mediocre or outright bad, but at any rate don’t rise to the threshold for acceptance at top venues.

I believe it is time to adopt a systematic approach to thinking about these stresses, and to use the insights gained to tailor a response. My belief is that an increasingly large percentage of our community is already

frustrated with the difficulty of finding outlets for their work, and this frustration is certain to grow. Rejected papers churn within the system, amplifying the underlying problem. The authors, perceiving the field as unfair, biased against them, and controlled by insiders, react in kind. Yet such trends bring us all dangerously close to improprieties such as duplicate parallel submissions, misrepresentation of authorship on multi-authored papers, and “politically inspired” reviewing. These are dangerous trends, and we mustn’t allow them to spiral out of control.

In what follows, I want to say more about the evidence that problems are arising (much of it anecdotal), and then suggest steps we might take to remedy them.

2. Evidence of a Problem

As noted, my contention is that we confront a variety of stresses running in two directions. Perhaps the more superficial direction is the sense of being overburdened that so many of us in the field are experiencing, with endless requests to participate on (or run) conference program committees, to read mountains of papers and proposals, and to somehow “track” a literature that is vast and seemingly expanding at an exponential pace. If one flogs the PC, they can still do an outstanding job (I did this with SOSP 05), but at what cost?

For those of us involved with program committees, particularly over long careers, I think the evidence of growing problems is hard to deny. Conferences are seeing larger and larger numbers of submissions – I remember the days when SOSP received 80 or fewer submissions. Today, we’re at two or three times as many. SIGCOMM gets more than 300 papers. And the trends suggest that the numbers are only going to grow, at least for a while into the future.

Overwhelmed by the huge numbers of submissions, most PCs have turned to multi-round processes in which the first-round reviews are farmed out, often to students who may do an erratic reviewing job. This drives a selection process that can whittle the initial set of submissions down to a more manageable size, but creates a serious signal-to-noise issue (particularly if paper rankings include these external reviews as well as internally generated ones).

We all know that some good papers die at this stage, but because plenty of good papers survive, the problem is hidden: the quality of our conferences isn't greatly impacted. As for the unlucky folks who lose out in the first round – we've all experienced that – well, one assumes that they will try again, and hopefully do better next time. Yet this situation is clearly unsatisfying because it implies at least some risk, perhaps substantial, of round-one unfairness. After all, as many as 2/3 of all papers will be rejected and in many cases, no PC member will even have glanced at the submission!

Complicating the situation is the human inability to read vast numbers of 12 or 14 page papers. Even the second round can be an immense load. For SOSP 05, some PC members (including me) read all the submissions, and wrote reviews for perhaps 40 or 50. This worked, but the physical toll of doing so was just enormous. Even the more common situation in which a PC member is asked to review “just” 20 or 30 papers is too much for many to handle. Overloaded, the PC starts to skim papers, reading only the “good ones” in any detail. But skimming is an error-prone way of reading dense technical work.

This creates a strange selection process, in which work that can't be described in a short paper is often never published at all, while work that can elegantly fit the format wins best paper awards as much for the relief the PC felt at reading something that made its points quickly and clearly as for the underlying merit of the work. Where is definitive paper on Windows, or .net, or Apple's operating system? We're increasingly trapped in a sound-bite world where ideas that just need longer expositions can't be published in conferences.

The second direction in which we're seeing signs of stress relates to the experiences all of us are having with good papers that get rejected in seemingly unfair ways:

- Who hasn't had papers that were rejected in the first round of reviews at a top conference, with just two reviews, one or both of which seemed almost completely clueless? Who hasn't expressed anger

at the system? Here are two little “factoids” to illustrate the depth of the issue: when I sent out the SOSP reviews, we discovered that in one case, a rejected paper had missed the initial cut *on the basis of a review that was clearly written about some other paper*. In the NSDI 2008 process, just before the PC meeting, we noticed that a few papers had no reviews at all – they had missed the cut because the “average score” was (obviously, under the circumstances) zero. One of those last-minute NSDI “catches” turned out to be in the top third of papers ultimately accepted by the conference.

- Most of us are learning to write papers in a manner calculated to appear to those beleaguered first-round reviewers. To get into SOSP or SIGCOMM a paper has to survive two thresholds: it must get past the two randomly selected students, and then must get past the six or so PC members who are most knowledgeable about the topic.
- Some subcommunities are increasingly bitter. The European community remains convinced that conferences discriminate against their work because of minor grammatical issues or similar writing problems. One tends to dismiss these concerns. Yet having read huge numbers of first-round reviews, I've realized that some external reviewers really can be bizarrely harsh, literally taunting authors for minor grammatical mistakes or other signs of their non-native writing technique – and often disparaging the scientific content of the work.
- More and more researchers confirm, when asked, that they routinely need to try many times before their papers are accepted. Many have started to generate small deltas on a basic idea as a way to submit the same work in parallel to multiple venues without overtly breaking the rules.
- Students report being under huge pressure to publish in the top venues, and are sending in unfinished work knowing that the whole game is something of a roulette wheel. If one of their papers gets in, they can always pull more material into the final version.
- Many teams are starting to generate papers with very long lists of authors. Obviously, this sometimes is appropriate, but it also makes sense as a remedy for a situation in which perfectly good papers often get rejected. After all, advisors have an obligation to ensure that their students graduate with reasonable CVs!

3. So, what should we do?

To fix these problems we need to fix our conferences. For reasons of brevity, let me just toss out a few ideas that, I believe, could have a big impact. In each case I'll also point out potential secondary consequences that my suggestion could trigger.

Reducing the “sound bite” paper phenomenon.

I've noted that a consequence of the overloaded PC situation is that all papers seem to be sound-bites. I think this is a bad thing, and that we can actually address both problems at the same time. Suppose that we were to eliminate page limits for our major conferences: a paper can be of any length. After all, we're publishing on the web these days; who cares about the page count? Indeed, one might argue that technology favors the opposite of length restrictions: papers should include color graphics, demos, videos or other materials, as appropriate – a paper should be a live document, not something that reduces to black and white bits on printer paper. The obvious rejoinder is that such a change would make the PC's job impossible. But this can be addressed using extended abstracts – let's say 6 pages or less. By taking this step, we open the door to papers on genuinely large systems that simply can't be described accurately in 14 pages, while preserving the obligation that the authors be able to communicate the innovative ideas in a brief form that either captures enough interest to motivate reviewers to “read the details”, or permits the PC to move on quickly without suffering through 25 pages of confusion. Everyone comes out ahead – and people who like to build really big systems have some chance of reporting their work in conference venues.

Con: Perhaps PC's will start to base their decisions entirely on the six-page abstracts, treating even a standard 14-page paper as “too long” to really read.

Exploiting social networks to improve reviewing

What about the problems of the first-round PC? It seems to me that our community could explore “social networking” mechanisms as a way to improve that first-round process dramatically, and also to regain the consistency we seem to have lost. The idea is to harness the thousands of researchers who regularly attend our major conferences. I'm imagining that we would create a web resource -- a kind of network of reviewers having some maturity in the field, perhaps a few publications of their own, and extensive exposure to the best work. Conference PC's would use this large resource in the

first round, in effect trusting our own traditional audience to a greater degree than we trust the random process by which a PC chair today assigns some paper to PC member X, who then randomly hands it to students Y and Z, producing completely random reviews from people who have never been a part of the community and who are naturally inclined to be overly critical and to overly favor work in their own areas of interest: our mature researchers have long since shed these flaws of youth.

Think of this as a kind of specialized search: given a submission, who would be the best non-conflicted reviewers in our “base” of candidate reviewers? Given that we could potentially marshal literally thousands of participants, we should think about using the same tools that enable search in the web to build automated paper assignment tools, automated conflict of interest detectors and automated reputation mechanisms. People who routinely refuse reviewing requests should be publicly blacklisted, as should PC members who accept the role and don't do the work: for those whose promotions may depend upon “professional service”, just tracking the statistics would be a powerful incentive to participate.

Much as the Internet ArXiv maintains a history of papers submitted, reviews and other commentary, and of later revisions, we could also consider creating a paper-tracking system that might span all our major conferences. If a paper is submitted to OSDI, rejected, and then revised and resubmitted to SOSP, why shouldn't the PC have a chance to see the prior history? Over time, one could imagine evolving a system in which papers would be submitted to “the field of networks”, or “of systems” and conferences could then choose among the best currently unpublished work. But even if we never move much beyond the opportunity for an author to receive criticism, respond, and have the history of that interaction preserved for later PCs to glance at, we would raise the quality level of the field substantially. Wiki pages could be used to permit a form of open community comment, perhaps offering chances both to notice that work is more incremental than the PC realized, or conversely that work is exciting broad interest when the PC hadn't noticed the key idea.

Con: Reputation systems are notoriously error prone. Blacklists might damage careers in profound ways. Merely having attended NSDI once or twice is no proof that an individual is at all competent in the field. Revised and resubmitted papers may be so different from the earlier version that reading the old reviews would bias the PC against a far improved paper.

Institutionalize the “rebuttal” opportunity

As noted, some conferences now offer a very brief rebuttal opportunity – they send out the reviews just before the PC meeting, and invite authors to respond. The goal is to avoid gross miscarriages, not really to encourage a lively debate. This, it seems to me, is a highly effective remedy to the risk that really confused reviews might pollute the decision process.

Cons: To be useful, a rebuttal needs to be very short, very pointed, and respectful in tone. If the rebuttal doesn't result in a review being discounted, a bad review might still taint the paper ranking. Rebuttals can trigger anger within the PC (particularly if the rebuttal attacks a review written by one of the PC members).

Increasing the number of outlets for research

Today, we're solving the “too many papers” problem by rejecting enormous numbers of papers. This is a harsh approach from the perspective of researchers who need to publish to get raises and retain their jobs, and may also be denying us exposure to whacky, out-of-the-box ideas that reviewers find hard to swallow. Yet in rejecting these oceans of papers, we create the very downstream problem that has us so overloaded! It seems to me that the right solution is to offer such work an outlet. Like medical conferences, many of our largest venues need to think about having short-paper tracks.

If NSDI or SOSP is going to receive 250 submissions, we may be right to continue to limit the conference to 25 or so full-length papers. But we could deliberately accept and publish an additional 25 “short papers”, using a WIPS format for the talks but including the full length papers in the proceedings, perhaps identified separately (“SOSP Short communications track”, for example). Doing so would open our doors to a much wider range of ideas without weakening the core conference. The WIPS track fails to accomplish this today because the corresponding papers aren't published and hence can't be cited; a WIPS talk reveals an idea and yet ensures that the authors will only get credit if they manage to publish that idea later in a full-length paper. This is an unfortunate dynamic, and sometimes harmful to the student. By elevating the WIPS track slightly, and giving a citable publication, we ensure that ideas are properly attributed and also that the audience intrigued by a 5-minute “short communications” presentation can read the 15 pages of details if they so desire. We can still keep the true WIPS session, of course; it

also serves a second role of exposing very early ideas, and we shouldn't abandon that goal.

This approach would have many benefits. True, publishing in the SOSP “communications” track may not be remotely as prestigious as publishing in the regular track, but it still would count as a publication. For those who can only use travel funds if they have a paper, we would open the door to participating in our conferences. And while researchers who feel they have a grand cru concept might not want to publish in this table-wine manner, those who grab at the solution will remove their papers (mostly solid but uninspiring work) from the mix, reducing load on PCs and freeing the PC members to focus on the stronger submissions.

Con: If we aren't careful about standards, nobody will read most of these short papers.

Level the playing field

Earlier, I mentioned that the current situation is fostering perception that the playing field isn't level. This is a very damaging problem and one we really need to deal with. It is vitally important that everyone have an equal chance.

I think there are many things we can and should do. The trend towards short rebuttal opportunities, as used by DSN and ECOOP, is worth adopting more broadly: if a reviewer does a very poor job, the author has a chance to point this out. One SOSP review turned out to be misfiled in the year I ran the conference: a review for paper X was uploaded for paper Y. We didn't catch this mistake, and while it probably made no difference, it was very unfair. A rebuttal opportunity would have saved the day.

Unfairness extends to many dimensions of the process. As a researcher interested in scale, I can't help but be disturbed by the tendency of PCs to demand types of experimentation that can only be undertaken by employees of the largest companies. Such companies have taken to writing papers that depend on s proprietary data sets or experimental infrastructures: “if you want to work on topic XYZ, you need to do it at Yahoo! (or Google, or Amazon, or MSN). Nobody can compete in this game except the big guys.

I think we can push back here: Papers in which work was evaluated using test sets that are not available to the public, or made available as attached secondary materials with the submission, should be treated especially harshly, and PCs should be instructed in the im-

portance of a level playing field that would give all researchers a chance to compete with new ideas in any space. To do otherwise is to tacitly accept that in order to do good research, one must work at Google or Amazon, and this (it seems to me) is not a role our field wants to play. Frankly, I think that if Google wants to publish papers about systems running on 10,000 nodes, they should have an obligation to do their work in a form that I, as a non-Google employee, can still validate and perhaps improve upon. Mere ownership of a lot of nodes shouldn't give the company some form of exclusive lock over an endless series of papers at SOSP, OSDI, NSDI and other top conferences.

Con: Companies may resist... although trends are promising: Yahoo! is releasing all sorts of data sets, and Google, Yahoo! and Amazon are creating big research clusters.

Push back on long author lists

Clearly, there are papers for which a long author list is completely appropriate. But we need to create disincentives that would serve as a counter-pressure against artificially long author lists. I think one could accomplish the goal easily by establishing a clear policy on what it means to "contribute in a significant way" to a result, and requiring that authors for papers with more than three or four co-authors simply attest, individually, that they did a fair share of the work. Perhaps I'm not cynical enough, but I personally doubt that students want to launch their careers in a dishonest way, and hence I believe that a student asked to attest that he or she really contributed to a paper will balk if, in fact, they contributed nothing of any consequence at all.

Con: It may be very hard to define significant contributions in a standard way. Legalistic-sounding definitions just invite people to act like lawyers and propose weird workarounds that comply with the letter of the law but in fact evade the intent.

4. So, what should we do?

Although this white paper was motivated by what I believe to be serious problems, I'm actually very optimistic that the field can address them. None of these problems represents a fundamental breakdown of ethics or research quality – the most foundational issue is really an issue of success, which is that the number of strong researchers in the community has never been larger, and the level of activity is rising. We simply need to learn to cope with a flood of solid work, not necessarily earthshaking in quality or implications, but

still worthy of capturing. If we can address this core need, we go a long way towards resolving problems that are becoming a huge annoyance not just to those of us who run the conferences, but to those who submit to them and attend them, too.

5. Acknowledgements

I want to thank the WOWCS program committee for their comments and suggestions; I can't any time a white paper of mine resulted in more pages of feedback than there were in the original submission itself! This revision is longer, but hopefully better structured and more balanced. I would also like to acknowledge the grants that support my research, from NSF, AFOSR, AFRL and Intel Corporation.