

Program Committee Overload in Systems

Ken Birman and Fred B. Schneider
Dept of Computer Science, Cornell University

Major conferences in the systems community—and increasingly in other areas of Computer Science—are overwhelmed by submissions. This could be a good sign, indicative of a large research community of researchers exploring a rich space of exciting problems. We’re concerned that it is instead symptomatic of a dramatic shift in the behavior of researchers in the systems community, and this behavior will stunt the impact of our work and retard evolution of the scientific enterprise. This essay explains the reasoning behind our concern, discusses the trends, and sketches possible responses. However, some problems defy simple solutions, and we suspect that this is one of them. So our primary goal is to initiate an informed debate and a community response.

1. The growing crisis

The organizers of SOSP, OSDI, NSDI, SIGCOMM¹, and other high-ranked systems conferences are struggling to review rapidly-growing numbers of submissions. Program committee (PC) members are overwhelmed. Good papers are being rejected on the basis of low-quality reviews. And arguably it is the more innovative papers that suffer, because they are time-consuming to read and understand, so they are the most likely to be either completely misunderstood or underappreciated by an increasingly error-prone process. These symptoms aren’t unique to systems, but this essay focuses on the systems area because culture, traditions, and values do differ across fields even within Computer Science—we are wary of speculating about research communities with which we are unfamiliar.

The sheer volume of submissions to top systems conferences is, in some ways, a consequence of success: as the number of researchers increases, so does the amount of research getting done. To have impact—on the field or the author’s career—this work needs to be published. Yet the number of high quality conferences cannot continue growing in proportion to the number of submissions and still promise presenters an influential audience, because there are limits on the number of conferences that researchers can attend. So attention by an ever-growing community necessarily remains focused on a small set of conferences.

The high volume of submissions is also triggering a second scaling problem: the shrinking pool of qualified and willing PC candidates. The same trends that are making the field exciting also bring all manner of opportunities to top researchers (who are

¹ ACM Symposium on Operating Systems Principles (SOSP), ACM-USENIX Symposium on Operating Systems, Design and Implementation (OSDI), ACM Symposium on Networked Systems Design and Implementation (NSDI); the Annual Conference of the Special Interest Group on Data Communication (SIGCOMM). This is a partial list and includes at most half of the high-prestige conferences in our field.

highly sought as PC members). Those who do serve on PCs rightly complain that they are overworked and unable to read all the submissions.

- If submissions are read by only a few PC members then there will be fewer broad discussions at PC meetings about the most exciting new research directions. Yet senior PC members often cite such dialog as their main incentive for service.
- If fewer senior researchers are present at the PC meeting then serving on the PC no longer provides informal opportunities for younger PC members to interact with senior ones.

And a growing sense that the process is broken has begun to reduce the prestige associated with serving on a PC. Service becomes more of a burden and less likely to help in career advancement. When serving on a PC becomes unattractive, a sort of death spiral is created.

In the past, journal publications were mandatory for promotions at the leading departments. Today, promotions can be justified with publications in top conferences (see, for example, the CRA guidelines on tenure²). Yet conference publications are shorter. This leads to more publications per researcher and per project, even though the aggregate scientific content of all these papers is likely the same (albeit with repetition for context-setting). So our current culture creates more units to review with a lower density of new ideas.

Conference publications are an excellent way to alert the community to a general line of inquiry, or to publicize an exciting recent result. Nevertheless, we believe that journal papers remain the better way to document significant pieces of systems research. For one thing, journals do not force the work to be fractured into 12-page units. For another, the review process, while potentially time-consuming, often leads to better science and a more useful publication. Perhaps it is time for the pendulum to swing back a bit.

2. Looking back and peering ahead

How did we get to this point? Historically, journals accepted longer papers and imposed a process involving multiple rounds of revision based on careful review. Publication decisions were made by standing boards of editors, who are independent and reflective. So journal papers were justifiably perceived as archival, definitive publications. And thus they were required for tenure and promotions.

This pattern shifted at least two decades ago, when the systems researchers themselves voted with their feet. Given the choice between writing a definitive journal paper about their last system (having already published a paper in a strong conference) versus building the next exciting system, systems researchers usually

² http://www.cra.org/reports/tenure_review.html

opted to build that next system. Computer Science departments couldn't face having their promising young leaders denied promotion over a lack of journal publications, so they educated their administrations about the unique culture of the systems area. With journal publication no longer central to career advancement, more and more researchers chose the path offering quicker turn-around, less dialog with reviewers, and that accepted smaller contributions (which are easier to devise and document).

As submissions declined, journals started to fill their pages by publishing material from top conferences. Simultaneously, under cost pressure, journals limited paper lengths, undercutting one of their advantages. Reviewers for journals receive little visibility or thanks for their efforts, so it is a task that often receives lower priority. And that leads to publication delays that some researchers argue make journal publication unattractive, although when ACM TOCS³ (a top systems journal) slashed reviewer delay, researchers still shunned submitting papers there.

Simultaneously, the top conferences have also evolved. Once, SOSP and SIGCOMM were self-policed: submissions were not blinded, so submitting immature work to be read by a program committee populated by the field's top researchers could tarnish your reputation. And the program committees read *all* the submissions, debating each acceptance decision (and many rejections) as a group. An author learned little about that debate, though, receiving only a few sentences of hastily written feedback with an acceptance or rejection decision.

Today, author names are hidden from the program committee, the top conferences provide authors of all submissions detailed reviews, and there are more top conferences (e.g., OSDI and NSDI) for an author to target. So authors feel emboldened to submit almost any paper to almost any conference, because acceptance will advance their research and career goals, but rejection does them virtually no harm. In fact, a new dynamic has evolved, where work is routinely submitted in rough, preliminary form under a mentality that favors a cycle of incremental improvements based on the detailed program committee feedback until the work exceeds the acceptance threshold of some PC. And often that threshold is reached before the work is fully refined. Thus, it is not uncommon to see publication of an initial paper containing a clever but poorly executed idea, a much improved follow-on paper published elsewhere, and then a series of incremental results being published. Perversely, this maximizes author visibility but harms the broader scientific enterprise.

Thus we see a confluence of factors that amplify—increasing the magnitude without adding content to a signal—the pool of submissions. Faced with huge numbers of papers, it is inevitable that the PC would grow larger, that reviewing would be done outside the core PC, or that each PC member would write reviews for only a few papers. The trend towards web-based PCs that don't actually meet begins to look sensible, because it enables ever-larger sets of reviewers to be employed without having to assemble for an actual meeting. Indeed, even in the face-to-face PC model,

³ ACM Transactions on Computer Systems (TOCS).

it is not uncommon for the PC meeting to devolve into a series of subgroup discussions, with paper after paper debated by just two or three participants while twenty others read their email.

Reviews written by non-PC members, perhaps even PhD students new to the field, bring a new set of problems. What does it mean when an external reviewer checks “clear accept” if he or she has read just 2 or 3 out of 200 submissions and knows little of the prior work? The quality rating of a paper is often submerged in a sea of random numbers. Yet lacking any alternative, PCs continue to use these numbers for ranking paper quality. Moreover, because authorship by a visible researcher is difficult to hide in a blinded submission (and such an author is better off not being anonymous), work by famous authors is less likely to experience this phenomenon, amplifying a perception of PC unfairness.

Faced with the painful reality of large numbers of submissions to evaluate, PC members focus on flaws in an effort to expeditiously narrow the field of papers on the table. Genuinely innovative papers that have issues, but could have been conditionally accepted, are all too often rejected in this climate of negativism. So the less ambitious, but well-executed work trumps what could have been the more exciting result.

Looking to the future, one might expect electronic publishing in its many manifestations to reshape conference proceedings and journal publications, with both positive and negative consequences. For example, longer papers can be easily accommodated in electronic forums, but authors who take advantage of this option may make less effort to communicate their findings efficiently. The author submits camera-ready material, reducing production delays, but the considerable value added by having a professional production and editing staff is simultaneously lost.

As the nature of research publication evolves, the community needs to contemplate two fundamental questions:

- What should be the nature of the review and revision process? How rigorous need it be for a given kind of publication venue? Should a dialog involving referees’ reviews and authors’ revisions plus rebuttals be required for all publication venues or just journals? How should promotion committees treat publication venues—like conferences—where acceptance is highly competitive but the decision process is less deliberative and nobody scrutinizes final versions of papers to confirm that issues were satisfactorily resolved? How do we grow a science where the definitive publications for important research are neither detailed nor carefully checked?
- Should we continue to have high-quality, “must-attend” conferences, with the excitement, simultaneity, and ad hoc in-the-halls discussions that these bring? If we do, and they remain few in number, does it make sense for these to be structured as a series of plenary sessions in which (only) the very best work is presented? As an alternative, conferences could make much greater use of

large poster sessions or “brief presentation” sessions, structured so that no credible submission is excluded (and printing associated full papers in the proceedings). By offering authors an early path to visibility, could these kinds of steps reduce pressure?

3. A high level view: What must change (and what must not)

An important role—if not *the* role—of conferences and journals is to communicate research results. *Impact is the real metric.* And in this we see some reason for hope, because a community seeking to maximize its impact would surely not pursue a strategy of publishing modest innovations rather than revolutionary ones. Force fields are needed to encourage researchers to maximize their impact, but creating these force fields will likely require changing our culture and values.

Another paper⁴ in this journal suggests a game-based formulation of the situation, where the winning strategy is one that incentivizes both authors and program committees to behave in ways that remedy the problems discussed above. One can easily conjure other characterizations of the situation and other means of redress. But any solution must be broad and flexible, since systems research is far from a static enterprise. A solution must accommodate a field that is becoming more interdisciplinary in some areas and more specialized in others, challenging the very definition of “systems”. For example, the systems research community is starting to embrace studying corporate infrastructure components that (realistically) can only be investigated in highly exclusive proprietary settings—publication and validation of results now brings new challenges.

Nevertheless, some initial steps to solving the field’s problems are evident. Why not make a deliberate effort to evaluate accomplishments in terms of impact? To the extent that we are a field of professionals who advance in our careers (or stall) on the basis of rigorous peer reviews, such a shift could have a dramatic effect. We need to learn to filter CVs inflated by the phenomena discussed earlier, and we need to publicize and apply appropriate standards in promotions, awards, and in who we hold-up as our leaders.

PCs need to adapt their behavior. Today, PCs are not only decision making bodies for paper acceptances but they have turned into rapid-response reviewing services for any and all. If authors of the bottom 2/3 of the submissions did not receive detailed reviews, then there would be less incentive for them to submit premature work. And even if they did submit half-baked papers, the workload of the PC would be substantially decreased given the reduced reviewing load. If some sort of reviewing service is needed by the field (beyond asking one’s research peers for their feedback on a draft), then rather than overloading our PCs, we should endeavor to create one—the web, social networks, and ad hoc cooperative enterprises like Wikipedia surely can be adapted to facilitate such a service.

⁴ Scaling the Academic Publication Process to Internet Scale. Jon Crowcroft, S. Keshav, Nick McKeown. Submission to CACM

Finally, authors need to revisit what they submit and where they submit it, being mindful of their obligation as scientists to help create an archival literature for the field. Early, unpolished work should be submitted to workshops or conference tracks specifically designed for cutting edge but less validated results. Presentation of work at such a workshop should not preclude later submitting a polished paper to a conference. And publishing papers at a conference should not block submitting a definitive work on that topic for careful review and ultimate publication in an archival journal.

Absent such steps or others that a community-wide discussion might yield, we shall find ourselves standing on the toes of our predecessors rather than on their shoulders. And we shall become less effective at solving the important problems that lie ahead, as systems become critical in the our society. Older and larger fields, such as medicine and physics, long ago confronted and resolved similar challenges. We are a much younger discipline, and we can overcome those problems too.

Acknowledgments. We are grateful to three CACM reviewers for their comments on our original submission. Jon Crowcroft, Robbert van Renesse, and Gün Sirer also provided extremely helpful feedback on an early draft. We are also grateful to the organizers and attendees of the 2008 NSDI Workshop on Organizing Workshops, Conferences and Symposia (WOWCS), at which many of the topics discussed here were raised.