Best Practices for the Care and Feeding of a Program Committee, and Other Thoughts on Conference Organization

Fred Douglis
IBM T.J. Watson Research Center

Abstract

I provide several lessons learned from running a number of conference program committees over the past decade, as well as some additional thoughts on conference organization and the reviewing process. Topics include how to deal with poor or absent reviewers, inbreeding among PC members, starting a new conference, and several other issues.

1. Background

As someone who has chaired a conference or workshop every 2-3 years for the past decade or so, when I came across the call for a workshop on organizing workshops, I had two reactions. One was that this was a fantastic idea and an opportunity to really share experiences, good and bad, to try and improve future events. The proceedings of such a workshop should be a veritable owner's manual to help train new conference organizers. (In fact, I was then approached to help create a new wiki to hold the collected "words of wisdom" resulting from the workshop: see:

https://wiki.usenix.org/bin/view/Main/Conference/Collected Wisdom or http://preview.tinyurl.com/26dchd

The other was that it could easily be a dismal failure because, while many have opinions about the topic, it could be that few would actually write about them. In retrospect, there were probably just enough submissions to ensure a lively set of discussions at the workshop. Hopefully the wiki will enable such discussions to continue past this one event and to open the discussion to other participants.

I begin this paper with some lessons learned from running program committees (PCs) and organizing new conferences (see the Appendix for a summary). I conclude with a few thoughts on some of the other topics raised in the workshop call. I believe my experiences apply to the "systems" conferences called out in the workshop call as well as to other venues; nothing I discuss is particular to systems venues.

Finally, an aside: this paper is by necessity a rather personal retrospective of issues I've encountered. Thus, I wrote in the first person more than I would in a technical paper. (Sorry, "it has far more of the first person

than a technical paper would.") If this bothers you, try going to the wiki and editing everything to be generic. It will read better in the long run, no doubt.

2. Lessons Learned

2.1. Those Who Forget History are Doomed to Repeat it

If I could impart just one piece of advice to a new program chair, it would be the importance of relying on the feedback of past chairs in evaluating PC members.

It seems that a good rule of thumb is that every PC will have at least one person who simply shirks his or her responsibilities, and fails to review the assigned papers. If a past chair can tell you someone to avoid, they are doing you a huge favor.

In my case, there are a couple of people who come to mind who were on a PC for me some time ago but failed to do their reviews. In both cases I am aware of the same thing happening sometime later, when I was (not surprisingly) never asked for my past experience with them.

The issue of identifying reviewers who have previously shirked their responsibilities is a delicate one. Informal discussions may identify specific cases but are unlikely to catch many offenders. A more organized method of tracking reviewer performance might be worthwhile but has some privacy considerations, as I discussed in a recent column [3].

This relates to one of the questions posed for this workshop: should we rate the reviewers? Ratings take two forms: quality of the reviews that are performed and general behavior of the reviewer. Despite arguments to the contrary [1], I do not believe that *authors* should rate reviewers, as I have seen at least one conference

do: for one thing, they have a very limited set of reviews to evaluate, so I know from personal experience that it can be hard to scale reviewers relative to the entire pool of reviews. In addition, there is the risk that an opinionated reviewer gets poor ratings simply because authors feel antagonized or disappointed.

On the other hand, program chairs and perhaps other PC members could rate reviews, much as some journals and magazines do, because they are exposed to a large enough pool to be well calibrated and because they are relatively impartial. These ratings should be useful in selecting future PCs and ensuring that reviews are a uniformly high standard. It is probably fair to assume that most reviews are "typical" and do not warrant special attention, but a mechanism for others who read a review to highlight that it is particularly good or bad could prove useful. We would not expect the program chair of a large conference to rate 1000 individual reviews.

One interesting tool in the old shell scripts used by USENIX program chairs in the late 1990s was to generate reports giving average reviewer scores and standard deviations. If people on the whole grade very leniently, or very harshly, or if they cluster all their scores around the average, the rest of the program committee should know this and weight those scores accordingly. Not all modern conference management systems make this information available, but they should.

2.2. Avoid Inbreeding

A number of conferences have a tendency to become rather inbred: they have a certain number of effectively permanent PC members, and only rotate a small fraction of their PC members from year to year. This is a **bad idea.** I believe that the core USENIX conferences, such as the annual conference and OSDI/NSDI[1] are pretty good in this regard, as are some other conferences like SOSP. Some other systems conferences retain a much higher fraction of PC members, which I

think results in a bit of tunnel vision, focusing on the same topics each year with much the same perceptions of what are good ideas and what are not.

Another possible aspect of inbreeding is the number of PC members from a particular organization or with a particular background. One USENIX security conference included a few people from one organization, and then the chair joined the same organization as the CFP came out, making it seem like he had selected 1/3 of the PC from his own organization. This looked bad to some, and while no one faults the chair for changing organizations,[2] there would not have been an issue of the other people didn't overlap so much. I can think of two other USENIX conferences that included over half the PC members with ties to the same department as the chair. I'm sure these PCs contained very talented people and I am not accusing them of bias; I am only suggesting that conferences need to avoid the appearance of being cliquish.

I think that conference organizers (such as USENIX) should establish guidelines for the number of PC members that can overlap in these respects, and then do a sanity check on PC lists prior to publishing the CFP. Some overlap with previous years is important, but too much overlap is terrible; finding that sweet spot would be a good topic for discussion at WOWCS. (I would recommend 20-30%.) Some conferences such as USENIX ATC have an informal policy of ensuring that a program chair serves on the PC the years before and after they chair it, which offers very strong continuity and should be adopted by all conferences.

One way to bring in new blood is to look at authors who have not previously served on the PC. When I chaired ATC'98, I took a USENIX bibliography to identify all authors of ATC or OSDI papers in the previous few years, then count their papers. I found a couple of people in my own department at AT&T who had published pretty much every year but never been on the PC … and sure enough they both turned me down, despite my pleas for the need for authors to play their part as reviewers.

A corollary to my point about identifying people who have published but not served is that I think it is, in

---

[1] A colleague reading this challenged me on this claim, expressing a belief that the turnover was much lower, so I did a quick experiment. I looked at the OSDI PCs from 2002-2008 and determined that it had exceptionally good turnover, with the vast majority of PC members being on just one PC, and only 2 (including one chair) appearing more than twice. NSDI was similar, with 59 people spread over 78 slots in 4 years.

[2] Another possible situation is when many people change jobs at once, so that by the time a conference occurs they overlap. This is unavoidable. I was focusing more on the situation where the PC started with a surprisingly large concentration in one organization.

general, a tragedy to appoint someone to a PC who has never published at a conference, if the conference has been around for at least a couple of iterations. Are there people who could serve on a PC for conference X based on their experience at conferences Y and Z? Sure. But if they haven't published at X, they either haven't been submitting there (meaning they may not be that interested in the conference and also that they may not be well calibrated to the material normally published there) or they've been having submissions rejected. There are generally enough published authors from previous conferences that these authors should be tapped.

There are certainly exceptions to the never-published-here guideline, such as an OSDI or NSDI author being invited to serve on the ATC PC. And there may be advantages to tapping a new community. But I believe there are risks as well. For instance, the first time I served on the Middleware PC, I had never actually been involved with the conference, but I arguably had the appropriate experience. On the one hand, this brought me into the community (I coauthored a paper there the following year and will be Industrial Track chair in 2008). On the other hand, the first year, I really was not yet calibrated to appreciate the types of content generally accepted to the conference, and my reviews may have been a bit more rigid than my colleagues'.

Bottom line: while some make the argument that it is useful to open a community to new people, I believe the right approach is to encourage the "outsiders" to submit to the conference first, and not jump right to the PC.

2.3. Keep a Short Leash

The first time I served on an ATC PC, for the 1997 conference, I was told that there were multiple deadlines: we should get 1/3 of our reviews in by the first deadline, another 1/3 by the next, and the rest by the end just before the PC meeting. Some conferences seem to follow this approach, and some don't. As a chair, I have usually used this approach and it has been a great help in identifying any PC members who need a little extra urging. When I haven't done it I have usually regretted it.

2.4. Be Explicit

As a program chair, you need to be as clear as possible with PC members about your expectations. To give a specific example, when I served on the OSDI'00 PC, I made the bad assumption that OSDI was more like the ATC than SOSP in the way the PC was run. I expected to review 20-30 papers, and then go to a PC meeting. Only as the papers came in did I find out the number of papers we were each expected to review was significantly higher, and there would be another round to provide more reviews of the papers that made the first cut.

Providing workloads and due dates when inviting PC members can go a long way toward ensuring that only people available to satisfy those demands will accept your invitation.

2.5. Watch for Conference Interactions

When I chaired USITS'99, we had to decide when to schedule the conference relative to other venues with which it might compete for papers. We decided to set the submission deadline to take place a couple of weeks after the notification for the ACM SIGCOMM conference, expecting that there might be high-quality papers that SIGCOMM would reject but which, after some modification, might be suitable to USITS. Note that we recognized that USITS was not as competitive as SIGCOMM; while we didn't want to fill our conference with SIGCOMM rejects, we thought there would likely be some appropriate submissions.

It turned out we needed to reschedule the conference and move its submission deadline to a couple of weeks before the SIGCOMM notification date. We did not want to lose out on those submissions, so with the permission of the SIGCOMM organizers, we arranged for people to submit to USITS even if their SIGCOMM submission was still under review. They checked a box indicating this overlap, so reviewers could hold off reviewing those papers until the SIGCOMM submission was resolved. If rejected, the authors had to provide us the SIGCOMM reviews: since they had no opportunity to revise the papers, and we had less time than usual to review, we wanted to be sure the SIGCOMM reviews were not "fatal." But to the contrary, my recollection is that we had two such submissions, both of which were rejected from SIGCOMM but with fairly good reviews, and both were accepted to USITS.

That tie-in was successful enough that I am rather surprised it is not more commonplace. It depends of course on several factors: a recognition that two conferences share content in common, a belief that the earlier conference is strong enough that even a rejected paper there might be worth consideration, and a willingness for the organizers of *both* conferences to accept a small overlap. The second submission must be

flagged to ensure that reviewers do not waste effort on it until the outcome of the first venue is resolved.

Regardless of whether small overlap in the submission windows might be permitted (and I acknowledge that it is far preferable to have some time to revise a rejected paper, but this timing is not always an option), the prospect of sharing reviews across conferences is appealing --- certainly when the conferences have some tie-in. In fact, it appears USENIX accomplishes this, in a manner of speaking, by encouraging program chairs to have overlap between conferences. When I last served on an ATC PC, there were a couple of people who'd served on the previous FAST and a couple who'd served on the previous OSDI, and when papers came up that had been rejected by those conferences, these PC members were able to share information about how the papers had fared. When an issue raised in an earlier conference was not dealt with in the revision, the paper was unlikely to be accepted. So … why not make the reviews available, rather than one or two reviewers? One would of course have to take the reviews of another venue with a grain of salt, in case there is bias, poor reviewing, or other issues.

Another factor in sharing reviews is that, like a magazine or journal submission that undergoes "major revision," there is a chance to identify what was problematic about a submission and how it was fixed – rather than starting every conference with a clean slate.

These issues are discussed in much greater detail in another WOWCS paper by Paul Francis [4].

3. Other Musings

3.1. Self-plagiarism

Many authors like to reuse text, but there are no hard and fast rules guiding what is appropriate. Copying background or related work verbatim is nominally a no-no but in practice is not a deal-breaker. On the other hand, reusing section after section is clearly a problem.

I have occasionally come across such cases of self-plagiarism, usually by accident. For instance, someone who serves on two PCs may see similar papers submitted to both. I would like to see a mechanized approach to detecting self-plagiarism with both published and submitted manuscripts, but there are numerous issues of privacy and intellectual property to deal with. Refer to another of my *Internet Computing* columns for additional discussion of this issue [2].

3.2. Rebuttals

I recently had my first experience with a conference (AAMAS'08) that offered the opportunity to rebut reviews prior to the final decision, and I think it was a *terrific* opportunity.

The rebuttal was limited to a small amount of text for each review, which I think is critical in making the process tractable. In our case, my coauthors and I identified one review in particular that made the comment that there was related work, but did not actually point us to that work; in addition, the overall recommendation was not favorable. Our rebuttal was merely a request for more specific information, as we were not aware of such related work. The review was not actually modified, but the paper was accepted. In the end I will never know if the rebuttal affected the decision.

Even if conferences do not permit rebuttals during the reviewing process, I think there is merit to having an author response, similar to the way journals deal with revised manuscripts. Currently this is done rather ad hoc, with authors contacting a program chair if they feel strongly enough about complaining, a situation that is extremely awkward. As a reviewer, though, I would like to be able to get feedback saying that I misunderstood something, or pointing out where I was not clear, and giving me the opportunity to respond.

3.3. Sponsorship

I will end with a cautionary tale.

WOWCS is sponsored by USENIX, a professional organization that manages conferences and ensures that the conference has the appropriate financial resources. If it loses money, USENIX makes up that money through other sources (such as bigger conferences). If something bad happens, USENIX has liability insurance to cover the conference and its organizers.

WCW, before it merged with the Workshop on Internet Applications to become the IEEE-sponsored Workshop on Hot Topics on the Web, was completely independent. When I chaired it, it was supposed to be held in Beijing and sponsored by some Asian institutions, but the SARS epidemic caused us to relocate it to the USA, and specifically my lab.

The workshop wound up losing money that year. But in fact there was no organization with "deep pockets" to cover its costs. IBM had agreed to host the workshop, not to fully sponsor it, but in the end IBM essen-

tially had to cover the bills; I wound up getting a last-minute conference sponsorship as a donation to the workshop, in exchange for IBM employees being able to come and go at will.

Worse, when we took a bus tour of New York City, had something happened, I'm sure it would not only be the tour operator on the receiving end of a lawsuit. An organization like USENIX or IEEE would have explicit insurance to cover this. IBM would be covered if someone sued the company, but what if they sued the organizers? I'm glad I didn't find out. As a reviewer pointed out, there are companies that offer insurance for specific events: these are worth investigating if one is not already covered by a professional organization.

3.4. Keeping a New Conference *on Track*

As chair of the IEEE Computer Society's Technical Committee on the Internet, I helped kick off SAINT in 2001. We tried to model this conference after the World Wide Web conference, complete with tracks, but we didn't properly estimate the reception a new conference would get, especially without having excellent publicity. We had a program committee of about 95 members (several tracks with 8-15 members per track) but only got 135 submissions. Oops.

One of the questions posed in the CFP was how to manage a large conference. So the first question, if the conference is new, is how large will it be? This can be hard to estimate. Guess too low and your PC will be swamped; guess too high and it's embarrassing. I think one rule of thumb for me is to never start a new conference expecting it to be as big as a comparable existing conference; we were dreaming. Once it gets large, though it's important to divide when needed. WWW waited at least one year too long to break into tracks (in 1999), since in 1998 I could identify a paper that was in my area of specialization, which I was never asked to review when on the PC, but which many people at the conference agreed had serious technical flaws. The wrong reviewers passed judgment on the paper. If a conference is broad enough that there will be many papers for which only a small subset of the PC will be qualified to review, then the conference probably should either be divided into tracks or disbanded as being too broad.

Given tracks within a conference program committee, there are a number of issues of fairness to consider. Not every track gets comparable quality submissions, so comparing papers or acceptance rates head-to-head may not be appropriate. But conference organizers need to decide up front whether they want to allocate space in the conference roughly by submissions (if a track gets X% of the submissions then X% of published papers come from that track), or based on overall quality (in which case a track might get a substantially higher fraction of its submissions accepted than the average). I personally favor the latter approach, which ensures a uniform quality standard for a conference regardless of how the submissions are distributed.

4. Summary

Running a conference is hard work. As an organizer, get all the help you can, especially from people who have run the same conference before. Make sure you have a formal organization standing behind the conference, to cover it financially and to provide appropriate insurance.

In forming a program committee, give the newcomers a chance, but be prepared for people to ditch their responsibilities. One way to prepare is to have some early deadlines from which you can recover early if a problem arises. And in the end, be sure that you know most of the PC yourself: friends tend to be more responsible than strangers.

Acknowledgments

References

1. Jon Crowcroft, S. Keshav, Nick McKeown, "Scaling Internet Research Publication Processes to Internet Scale," *WOWCS* 2008.

2. Fred Douglis, "All Systems Go: Collective Wisdom: A Modest Proposal to Improve Peer Review, Part 1," *IEEE Internet Computing*, vol. 11, no. 5, pp. 3-6, Sep/Oct, 2007.

3. Fred Douglis, "All Systems Go: Collective Wisdom: A Modest Proposal to Improve Peer Review, Part 2," *IEEE Internet Computing*, vol. 11, no. 6, pp. 3-5, Nov/Dec, 2007.

4. Paul Francis, "Thoughts on How to Improve Reviews," *WOWCS* 2008.

A. Experience

One of the reviewers commented that my list of conferences with which I had experience was too self-congratulatory. I was not sure whether to omit it or relegate it to an appendix, but here it is:

The conferences I have program chaired or co-chaired range from two USENIX conferences (the 1998 Annual Technical Conference (ATC) and 1999 USENIX Symposium on Internet Technologies and Systems (USITS), the precursor to NSDI) to a new IEEE Conference (the 2001 Symposium on Applications and the Internet (SAINT)) to two web venues, the 2003 Web Caching and Content Distribution Workshop (still known as WCW) and the 2005 World Wide Web (WWW) conference.   I was also a program vice-chair three times for WWW.

In addition to serving as program chair, I have been involved in the organization of numerous other conferences, either on the steering committee or, in one case, as general chair (the 4th WWOS, the precursor to HOT-OS).

Most of these conferences went pretty well. Some did not.