

Message from the USENIX ATC '20 Program Co-Chairs

1. Preamble

Every year, conference chairs share their thoughts about the conference. This is often chock-full of statistics about acceptance rates, etc. We'll include stats for sure. We felt, however, that we'd like to share with all of you some of the "inside scoop" of how a conference is organized and run from the Co-Chairs' perspective. Many authors don't know what happens behind the scenes; and even most PC members don't get to chair a conference of this magnitude. We are honored to have been selected to co-chair USENIX ATC '20. Running a conference like ATC is substantially more work than some other, even well established conferences, for the reasons outlined below (and that effort skyrocketed when the pandemic hit). There is a lot to know and we hope that this information will be valuable to all of you in the future.

2. Early Decisions

ATC is a large and complex conference to run for several reasons.

First, while it is a systems conference, ATC is fairly broad. Whereas conferences such as OSDI, FAST, NSDI, etc. see more focused papers, ATC has a much broader set of papers that span many systems conferences: networking, storage, operating systems, security, etc. This breadth meant that the PC's makeup had to be carefully balanced to ensure appropriate representation for all likely topics.

Second, ATC is also the place where new systems papers are often submitted for topics that may not have a home of their own. FAST, NSDI, USENIX Security, and others—all started in part because ATC was getting too many papers in those areas, suggesting it was time to form a new, more focused conference. In recent years ATC started getting papers on edge computing and applied machine learning; and this year we expected we might get for the first time ever papers in quantum computing (and we were right). This meant that the PC's makeup also had to include people who could review papers in newer and emerging topics such as machine learning and quantum computing.

Third, ATC receives submissions from many first time or relatively junior authors. Everyone has to start somewhere. But this meant that the PC had to do more work in reviewing and shepherding papers.

It would be nearly impossible to find ATC PC Chairs who are experts in all areas of systems research, including emerging ones. Recognizing this, we decided, for the first time, to recruit a few "PC Leaders" early on. We recruited five excellent PC leaders who, in addition to reviewing papers, would also help us form the PC itself. These five Leaders are experts in areas that us Co-Chairs were not as knowledgeable in, and they were instrumental in recommending and helping us pick members for this large PC.

Another important decision we made followed previous ATCs: we selected two "Submission Co-Chairs," one each at our respective institutions: Ketan Bhardwaj (Georgia Institute of Technology) and Dongyoon Lee (Stony Brook University). They were instrumental in assisting us run the conference submissions site (HotCRP) and overall, during the review process. They helped us analyze data, collect various statistics, write scripts, monitor online discussions, helped us run the (virtual) PC, and more. We cannot imagine running this conference without them, especially when the pandemic demanded a lot more effort on everyone's part.

2.1. Decisions we made early on: R1 rejections, topics/themes

Every chair gets a chance to redefine the process of running the conference somewhat. We looked at past conferences as well as past ATCs and decided to try a few changes.

2.1.1. Two review rounds and early rejections

ATC and other conferences have two rounds of reviewing. In the first round (R1), every paper gets a few reviews, and we decided to keep it at 3 reviews. There would then be online discussion and filtering. Those papers deemed worthy would advance to the second round of reviewing (R2) and receive about 2 more reviews. We aimed for 3 reviews in R1 and 5 in R2. There's an "art" of sorts to deciding how many reviews to have in each round: if you have too few, less informed decisions are made. But ironically, if you have too many reviews, in many cases that does not lead to better decisions. We have seen it before:

when you assign a paper more reviews, there's sometimes less convergence to an accept/reject decision; often the result is more variance, so converging on a decision is harder. No, more reviews aren't necessarily better. Rather, PC members have to be given time to discuss their reviews and try to convince each other of the merits of one's opinions.

Traditionally, papers that don't make it to R2 are slated for rejection. But as there's a second round of reviewing and a PC meeting, there's often a month or more between when R2 reviewing begins and all rejection notices are sent.

So we decided to send rejection notices for papers in R1 who didn't make it to R2, as quickly as we could. That was almost a month before final paper decisions were released. We did this to help authors, who now have even more time to consider their papers and the reviews they received, and decide how to revise the work for an eventual resubmission.

While sending R1 rejections helped authors, it meant more work for the PC and Co-Chairs, and a tighter schedule. Before any reviews are sent out, we decided to follow past models and perform a "Review Sufficiency Check" (RSC). In this stage, a paper reviewer is assigned to read all of the reviews and to ensure that they are complete, detailed, fair, respectful, proofread, and not missing useful information. An example of what an RSC check catches is when a reviewer might say "this work was done before," during this RSC phase, we would catch this and ask the reviewer to provide more details such as an actual citation to the "done before" work.

2.1.2. Paper topics

When authors submit papers, they check several boxes to indicate what topics their paper fits into (e.g., networking, storage, operating systems). Similarly, before PC members begin to review any paper, they log in to the conference reviewing system, HotCRP (yes, pronounced hot-crap :-), and configure their expertise level in each of the topics. PC members mark their expertise on a five-point scale from "no expertise," to "neutral," to "expert." This indicates the PC member's preference in reviewing or not reviewing papers on certain topics.

Assigning hundreds of papers to over 100 PC members is a daunting task. Thankfully, HotCRP has a built-in "auto-assignment" feature. HotCRP can match papers to PC members by mutual interest in a given topic. HotCRP's initial assignment is very helpful but is not perfect, so it still needs a lot of review and tweaking by the Co-Chairs.

One trend in recent years has been to increase the number of topics and even add "cross-cutting" themes (sometimes called "aspects"). For example, instead of just saying "networking," you break it down into local-area, wide-area, mobile, protocols, data-center, etc., networking. And instead of saying "storage" you break it down to file systems, NVMs, Flash storage, network storage, KV stores, etc. And then you add orthogonal themes that can be applied to any of the finer-grained topics, for example: performance, scalability, security, availability, reliability, scheduling, etc.

There was a good reason to add more and more fine-grained topics and themes: HotCRP could do a better matching of papers to reviewers, ensuring that the best-qualified experts would get to review a paper. But there was a growing problem with the proliferation of topics and themes.

First, asking PC members and authors to mark 60+ topics and themes was a growing burden.

Second, authors don't always mark their papers correctly, often missing important topics/themes.

Third, some PC members don't want to mark their topic preference on purpose: they would like to get a "random assignment" of papers from the pool of submissions. There is a valid reason for this. A good paper should not just appeal to experts in the paper's field, but to a wider audience from the same community. So by reviewing some papers by highly technical PC members, but not necessarily experts, you ensure the selection of papers with greater appeal to the systems community.

Fourth, and most important. We found that the "cross-cutting themes" were not helping much; in fact, they were likely hurting HotCRP's initial assignment of papers to reviews. Suppose an author marks their paper with one or more topics and the theme of "performance." A person who can review a "storage performance" paper well probably won't be able to review a "network performance" paper as well. Similarly, "scheduling" in data centers is very different from I/O scheduling or network-packet scheduling. Analyzing these themes, we decided that they were not helpful and even counter-productive, and so we decided to eliminate all cross-cutting themes from the USENIX ATC '20 CFP.

Next, we tackled proliferation in the number of themes. After much discussion that involved our PC Leaders and Submission Co-Chairs, we decided to narrow the list of topics from several dozen to just 14. The reason to reduce the number of topics was to lower the burden on authors and PC members when marking their expertise; but at the same time, we wanted to ensure that we would still be able to assign papers to expert reviewers. So we chose multi-topic themes such that expertise in one

topic is likely to translate to another. For example, one topic was “Distributed Systems, Clouds, Clusters, Data Centers.” The idea was that if a PC member can review a paper in “Distributed systems,” then they should also be able to review a paper on “data centers” reasonably well. No paper-to-reviewers assignment is ever perfect, but we felt that the way we chose themes and fewer of them helped reduce overall burden and still keep the reviewing quality high.

3. PC Selection Process

PC selection is a multi-dimensional optimization problem. Co-Chairs have to ensure that the PC is well balanced as much as possible across multiple dimensions (in no particular order):

- 1. Junior vs. senior people:** it’s good to have some “fresh blood” and first time PC members, but also to have senior people who provide the necessary “institutional memory” and the wisdom that comes with... gray hair.
- 2. Industry vs. academia:** it’s not enough to have just academics on a PC. We have to have representation from industry, who work on many real-world cutting-edge problems.
- 3. Domestic vs. international:** the systems community is spread around the world. We want representation not just from U.S. researchers, but from Europe, Far and Middle East, and more.
- 4. Male vs. female:** notwithstanding that genders are not binary, computer science and systems research are sorely lacking in female representation. This is true in academia (faculty and students), as well as industry. Inclusivity and diversity has proven to improve outcomes of many decision-making processes. Alas, our community has far fewer female researchers than we’d like to have. Many female researchers get asked repeatedly to serve on PCs, creating perhaps an unfair load. We tried to invite as many female PC members to join as we could: 24% of invited were female; and in the end our PC had 23% female representation.
- 5. Topic diversity:** it’s important for the PC to have just the right number of people to review papers in various topics. We reviewed stats from previous years, such as how many papers were received in traditional topics (e.g., storage, clouds) as well as emerging topics (e.g., edge, ML). We estimated how many papers we might get overall and how many papers in each topic. We planned for getting anywhere from 300 to 500 (worst case) submissions. In the end, we had just under 350 reviewed submissions. For each PC member we considered inviting, we had a spreadsheet to mark their expertise in various areas.

Next, we chose to follow previous year’s models and invite PC members in multiple rounds and for multiple effort levels. During the fall of 2019 we sent out eight different rounds of invitations. After waiting to hear back, we reviewed the current makeup of the PC along the five aforementioned dimensions. We used google sheets and docs extensively. We then adjusted our strategy and sent out the next round of invitations, consulting with our PC Leaders as necessary. Adjustment was necessary to ensure that, if we were short in one dimension above, that we doubled our efforts to catch up along that dimension without hurting other dimensions.

We offered three tiers of reviewing load. Heavy PC members were expected to review 16–20 papers; Light PC members were expected to review 8–10 papers; and External Review Committee (ERC) members were expected to review 1–3 papers at most. Heavy PC members were required to attend the physical PC meeting (which was converted to a virtual PC—more on that later). ERC members were selected for specialized expertise in narrow areas where we felt we would need a couple more reviews. All PC members were expected to review papers in two rounds and participate in both rounds’ online discussions. We first invited Heavy, then Light, then ERC. In some cases, those who declined our invitation to be Heavy PC members were offered to be a Light PC member (several agreed).

We made a total of 214 PC invitations. 42% of all invited declined (many are understandably busy). Of the 125 that accepted, we had 67 Heavy PC members, 48 Light, and 10 ERC.

CFP updated and posted, PC formed, HotCRP configured, we were ready...

... or so we thought.

Con permiso, Capitán. The hall is rented, the orchestra engaged. It’s now time to see if you can dance.

–Q to Captain Picard, “Q Who,” Star Trek: The Next Generation

4. Before Paper Submission Due Date

Anonymizing papers helps ensure the integrity of the conference. When neither authors nor PC members know each other's identity, this is called a double-blind review process.

We received dozens of emails from authors who asked for clarifications on how to anonymize their papers. This is always a difficult subject. If you anonymize your paper too much, you may be hiding information about useful prior work you did, which could help the current paper under submission. Ironically, if you anonymize your past work too well, you could be incorrectly accused of reproducing or even plagiarizing “someone else's work” where it's really your own work.

If you don't anonymize, or poorly anonymize, you expose your identity, even indirectly to PC members. There's no consensus whether it helps or hurts a paper's chances, but there is consensus that it biases the process one way or another.

Many ATC authors are first-timers or relatively junior, which explains why we received many such queries about the anonymization process. Our advice to authors was as follows. PC members are encouraged not to search the Internet for information that might reveal the authors' identities. So authors had to ensure that their papers were anonymized and self-contained. Citations to one's own paper should be stated in the third person: you can refer to your own past work as “John Doe et al. [1] did this or that” and then state how you improved on those “other” people's work.

There are times when anonymization can get very difficult. Suppose the authors are from a very large cloud operator and are reporting on their experiences running a million computers across dozens of data centers worldwide. This kind of experience is invaluable to report to the systems community, as most researchers can only dream of having access to such a large set of computers and its associated data sets. Therefore it doesn't make sense in such a paper, to “hide” details as to the size and scope of the environment being studied: hiding such details would likely hurt such a paper considerably. However, anyone reviewing such an anonymized paper would have a pretty good guess which company the authors work for. After all, we can all count on one hand the number of cloud operators with such a large operation.

The above was just one example. We generally told authors that anonymizing is a “best-effort” process: try your best to hide your identity but avoid going too far that it genuinely hurts your paper.

5. After Submissions Were Received

HotCRP had 408 records of submissions, but many were just from novice authors experimenting with HotCRP to see how it works; some authors re-uploaded their paper multiple times, creating duplicate submissions; and some authors registered the paper days earlier, but withdrew or never completed their submission.

There were just over 350 complete submissions. We inspected each one of them! Not surprisingly, we found over a dozen submissions that did not follow the rules: many did not anonymize and even had the authors' full names on the title page; some had bad fonts or bad PDFs; some had odd background colors or a watermark; some violated the formatting guidelines (fonts, sizes, margins, or the number of pages). We contacted each of those papers' authors and gave them 24 hours to fix their submission or we might have been forced to withdraw their papers. Most authors complied quickly. A couple of papers were withdrawn. We were left with 348 submissions that were ready to be reviewed.

6. Conflict Management and Integrity

A conference's reputation and prestige depend heavily on at least two important factors. First, the number, length, and quality of the reviews received. This is especially important for rejected papers. Getting a paper rejected is always disappointing; a good conference will provide reviews that would show authors that reviewers understood their paper, read it carefully, and provided constructive criticism to help the authors improve the work for a future resubmission.

Second, the perceived fairness of the process. In a conference like ATC, where over 80% of all papers are rejected, it is vital that the authors of all those rejected papers would feel that the process was fair and unbiased: that the decision on papers which got accepted or rejected was based on merit alone, and not on who reviewed, who was on the PC, or who are the conference organizers.

In 2019, there were stories circulating about possible poor practices in certain systems conferences. In one case this led to a tragedy. Organizations such as ACM and IEEE are still investigating various reports. And we have also heard from other senior members of our community, directly and indirectly, about possible conflicts and violations of anonymity rules.

We expected and followed three common conflict and anonymity rules:

1. PC chairs are prohibited from submitting papers to their conference. When recruiting our Submission Co-Chairs, we told them they would also not be allowed to submit papers to ATC'20. That is because all four of us had “superuser” privileges on HotCRP, allowing us to see any review.
2. Authors had to mark every PC conflict (and select a reason) for their paper. HotCRP is excellent at ensuring that conflicted PC members had no direct or indirect knowledge about papers those PC members are conflicted with: they can't see reviews, who reviewed, or scores; and they would be excluded from all online and PC discussions regarding conflicted papers.
3. Both of us Co-Chairs had papers marked as conflicts (often from our collaborators or ex-students). That's unavoidable. Thankfully there were no papers in which we were both conflicted. Therefore, we decided that each one of us PC Co-Chairs would alone handle all matters (e.g., PC discussions and decisions) relating to papers we were conflicted on; this included getting us PC Co-Chairs (or Submission Co-Chairs) excluded from parts of the live PC meeting discussions, as well as the selection of paper awards.

Nevertheless, we decided to “take it up a notch”:

First, we added a new statement to the CFP that read as follows:

Authors and others are prohibited from directly or indirectly communicating with any ATC '20 PC/ERC member about any potentially submitted paper. All requests should be made exclusively to atc20chairs@usenix.org. Violations of these guidelines may seek remedies as stipulated in the USENIX Conference Submissions Policy.

This was to ensure that PC members and authors should never communicate directly, regardless of whether a paper was accepted or rejected, even after the conference is long passed.

Second, our Submission Co-Chairs leveraged past scripts and data-sets that scour the Web, DLBP paper databases, and other resources. Those scripts attempt to identify additional or incorrectly marked conflicts for every paper. The scripts, for example, attempt to find if person A collaborated with person B on a recently published paper, person A submitted to USENIX ATC '20, person B was a PC member, but person B was not marked as a conflict for person A's submission. The scripts also looked at possible institutional conflicts, and more.

These scripts were far from perfect. They produced many false positives due to very similar names as well as their inability to perfectly match people's names with abbreviated initials. Our Submission Co-Chairs manually poured over thousands(!) of flagged potential conflicts from these scripts and together we narrowed them down to about 242 possible conflicts across 92 papers. We wrote separate scripts that emailed all authors and asked them to verify whether these possible conflicts were valid or not. Over 50% of all responses indicated that indeed these were valid conflicts for those papers, and so we marked additional conflicts.

We were now ready in earnest to assign papers to reviewers and begin the first round of reviews.

Phew.

7. Next Stages of Running a Conference

At this stage of any conference, about half of the Co-Chairs' work is done: forming a PC, posting a CFP, getting and sanitizing papers, and assigning reviews. We could now sit back and relax for a while...

... or so we thought.

COVID-19 happened. In other words, the Borg was unleashed on us before anyone was ready.

For many of us, the amount of work in our day-jobs increased considerably. Food and supply shortages. Dire predictions every day. No one knows what happens next. Many of us who are parents were having to juggle twice the workload and now having to provide full-time childcare, with schools and childcare facilities closed. People losing their jobs in droves. The economy tanking. The stress levels everyone was going through were through the roof.

We began to receive reports from PC members who were unable to meet the already tight review deadlines. Some reviewers reported stress, work and family problems, and even health problems related to COVID-19. Not prying, we tried our best to work with everyone's abilities and schedules. In some cases, we had to reassign papers and reviews to ensure we kept the expected review quality of ATC. This was a non-stop daily effort from mid-February to mid-April of 2020.

8. The PC Meeting

An in-person PC meeting is important. No amount of written reviews and numeric scores can replace the human experience. When people discuss a paper face to face, everyone can better gauge exactly how every reviewer feels about a paper. Seeing people's faces, hearing the inflection in their voices, and noting their body language are all vital cues that help us communicate better with one another. This is especially important because the PC meeting is often reserved for the most challenging borderline papers whose fate depends on informed discussions.

We had grand plans to hold a day-long physical PC meeting at Georgia Tech. The plan was to review no more than 70 papers during the PC meeting.

We started to get many emails from PC members who were concerned about traveling due to their own health, or their employer restricted or prohibited all travel. It was clear that holding a physical conference was not possible. So we canceled it and informed all PC members that it would be a virtual PC.

We had no idea how we would run a virtual PC yet. But we realized that we would not be able to hold a day-long virtual PC, nor would we have time to review ~70 papers virtually.

Based on past experience, we had already decided to pre-accept and pre-reject some papers during the R2 discussion period (and recall we also sent out early rejections for papers that were rejected in R1). The thinking was that if the PC can reach accept/reject decisions on some papers during the R2 online discussion period, and ahead of the PC meeting, then there would be fewer papers that had to be discussed at the actual PC. Past experience suggested that such papers' decisions rarely change during a PC meeting, so it was just going to waste time if we discussed them. Therefore, we assigned a discussion leader to each paper and asked them to see if a consensus could be reached during the R2 discussion period: pre-accept, pre-reject, or "need to be discussed at the PC." We gave PC members wide latitude to opine on those pre-accepted and pre-rejected papers. If any reviewer felt the decision was incorrect, then that paper would be discussed at the PC instead.

In normal times, it's just too easy to "punt" on making difficult decisions and merely push the paper to be discussed at the PC meeting. But now, having been forced to run a shorter, virtual PC, we could not afford to leave too many papers for the PC meeting itself. So we pushed our PC to make the difficult decisions early on, try to reach consensus, and pre-accept or pre-reject as many papers as possible. In order to provide for some cross-paper "calibration," we configured HotCRP so as to make all reviews and discussions visible to all non-conflicted PC members. This was difficult for everyone, especially during a worldwide pandemic. But our PC did admirably well. We were left with 38 papers to discuss at the PC meeting.

8.1. The Virtual PC (vPC)

Running a virtual PC (vPC) was a new thing. Few have ever attempted it. We could write ten more pages of "Chair Notes" about our experience. Wait, we did! You can find our full-length report on how we planned and ran our vPC here: <https://www3.cs.stonybrook.edu/~ezk/vPC.html>. So we'll only give you the highlights below:

1. We (two Co-Chairs and two Submission Chairs) evaluated multiple solutions and in the end, we chose Zoom as the most suitable one.
2. We shortened the PC meeting to just five hours. But we underestimated how long it would take. Indeed, we ran about 2 hours longer.
3. We polled our PC members for their schedules and current time-zones. And we clustered the papers to discuss into groups that best optimize for PC members' availability.
4. In a physical PC meeting, PC members who are conflicted are asked to leave the room, then they're called back in. With a vPC, we used Zoom's "Waiting Room" feature: it neatly allowed us to virtually move conflicted PC members from the actual PC to the waiting room, where their video and audio are disabled, they cannot hear or see any of the discussions, until they are let back in.

Given the unprecedented circumstances we faced, we believe the vPC went very well. And we strongly believe that USENIX ATC '20 maintains the high standards of quality despite all that everyone had to go through.

8.2. The Stats

And now, for the customary statistics from the paper review process.

Finalizing the program decisions for USENIX ATC '20 took 1,379 reviews, 4,193 discussion comments posted during the two online discussion rounds (post R1 and post rebuttals). The PC wrote a total of over 1.15 million words. Out of the initial 348 submissions, 150 papers (43%) were advanced to the second round, the remaining 198 papers received an early (R1) rejection notification. During the post-rebuttal online discussion period, the PC decided to pre-accept 42 highly-ranked papers and to pre-reject 70 more papers. The remaining 38 papers were further discussed during the vPC meeting. 23 of these papers were accepted and 15 were rejected. The final acceptance rate for ATC '20 is 18.7% (65 out of 348).

Among the accepted papers, 61 are regular full-length papers, 3 are short papers which were submitted as short, one paper will appear at the conference as a short paper but was originally submitted as a full-length 11-page paper, and another paper submitted as long was accepted as a short paper but given 9 pages. All accepted papers were initially conditionally accepted and assigned 1 or 2 shepherds, who ensured that the final version of the paper addressed the reviewers' feedback.

9. The Conference

With the paper selection process complete, we were almost done—just to organize the papers in sessions that fit the more-or-less standard schedule template that USENIX was going to provide...

... or so we thought.

Shortly after the vPC concluded and all decisions were communicated, unsurprisingly, it was determined that the actual conference will also have to be held as a virtual event. So, we were back in uncharted territory, trying to make the best decisions on how to organize and run a multi-day, multi-track USENIX ATC '20 conference—and for the first time fully online.

We spoke with Co-Chairs of other major systems conferences that switched to virtual mode—ASPLOS '20, Eurosys '20, and ISCA '20. We polled some of our PC members for their feedback from virtual conferences they had attended. We read blogs and recommendations on different virtual conference tools and experiences. We consulted on a weekly basis with the USENIX staff and the Co-Chairs of the workshops affiliated with ATC (HotStorage and HotCloud). And we mapped out the time zones of the authors of the accepted papers, and much more.

In the end, we arrived at a program that retained the original conference dates, organized in 12 technical papers sessions presented in two parallel tracks, and two plenary keynote addresses. With a goal of creating a balance among giving to authors an opportunity to present their work in front of a live audience, providing conference attendees with high-quality technical content, and shielding everyone from “Zoom fatigue,” we opted for a schedule that combines longer, asynchronously-delivered, pre-recorded presentations, and live sessions with shorter video presentations and Q&A.

The two keynotes are scheduled at the end of the first two days of the conference, with ample time for discussions and Q&A. The first keynote is by Professor Ethan Miller from UC Santa Cruz on “The Future of the Past: Challenges in Archival Storage” and discusses, for example, future storage technologies such as DNA and Glass. The second keynote is by Professor Margo Seltzer from the University of British Columbia, titled “The Fine Line between Bold and Fringe Lunatic.” Margo's talk hopefully sets a new tradition at USENIX ATC where the previous year's Lifetime Achievement Award Winner delivers one of the keynote addresses in the following year.

We look forward to three days of technical papers, keynote presentations, and discussions.

10. In Closing

Despite the tremendous amount of (unanticipated) work, we are thrilled to have had the honor of chairing USENIX ATC. We are thankful to everyone who helped and contributed along the way: to the authors who submitted their high-quality work to ATC, to the dedicated program committee and external reviewers who evaluated hundreds of submissions and provided constructive feedback to the authors, to the PC Leaders, Submission Co-Chairs and to the USENIX staff who provided invaluable advice and support.

We are excited to welcome everyone at the first-ever virtual USENIX Annual Technical Conference, and we hope you will enjoy it.

USENIX ATC '20 Program Co-Chairs
Erez Zadok, *Stony Brook University*
Ada Gavrilovska, *Georgia Institute of Technology*