

MAY 21-23, 2018 AT THE HYATT REGENCY, SAN FRANCISCO, CA

39th IEEE Symposium on Security and Privacy

Sponsored by the IEEE Computer Society Technical Committee on Security and Privacy
in cooperation with the International Association for Cryptologic Research

PC Chair Report on the IEEE S&P 2018 Submission & Review Process

Bryan Parno

CMU

Christopher Kruegel

UCSB

Overview

- Process changes
- PC survey
- Author survey
- Student PC report

Process changes

- Bulk submission rule: Removed
- Monthly review deadlines
- New decision option: Revise

New Reviewing Process

- Motivation:
 - Permit submission and acceptance of results any time they're ready
 - Frequent submission opportunities
 - Rapid turnaround to published paper
 - Reduce incentives to “cram” before deadlines and submit half-baked work
 - Improve quality of papers and review process
 - Start with papers written under less time pressure
 - Facilitate a more collaborative process between reviewers and authors
 - Smooth the reviewing workload to a low-volume, ongoing effort

New Reviewing Process: Rough Schedule

- Month X, Day 1

Submission deadline

~32
days

- Month X, Day 1-5
- Month X + 1, Day 7

Reviewer bidding (semi-auto via ML)

Reviews due

~13
days



- Month X + 1, Day 7-20
- Month X + 1, Day 20

Online discussion and more reviews

Virtual PC meeting via Zoom

Appearing at this Symposium

All new papers accepted by Feb. 1 or revisions accepted by Mar. 1

- Month X + 2, Day ~7

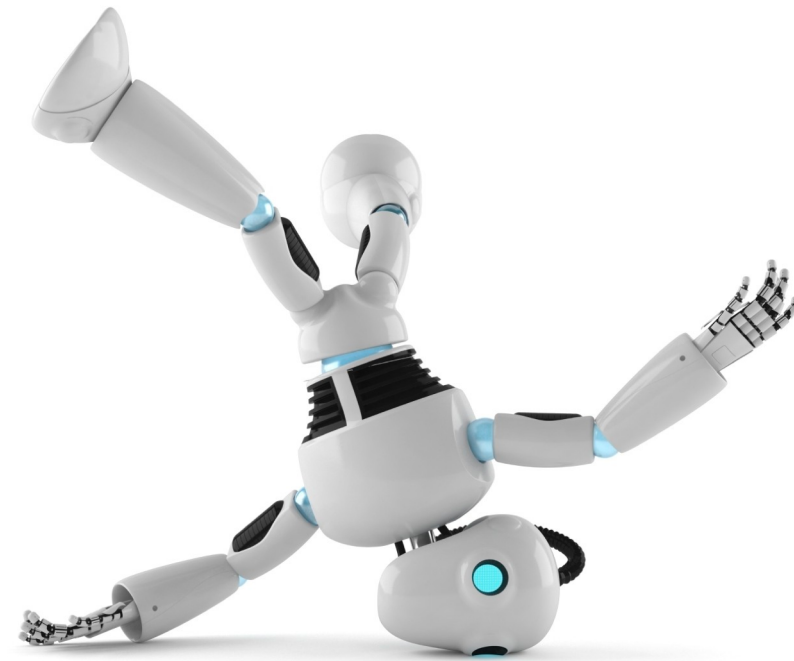
Accepted papers online (open access)

Semi-Automated Bidding

- Goals
 - Reduce PC's monthly review work
 - Shorten time to paper assignments
- Approach
 - Automatically scrape PC members' publication pages
 - Find PDFs, extract text
 - Compute bid based on similarity with each submission & enter in HotCRP
 - v1: Based on TPMS matching algorithm
 - Used at CCS for several years
 - v2: Based on latent Dirichlet allocation (LDA)
 - v3: Based on StarSpace (a general-purpose neural model for efficient learning of entity embeddings)
 - v4: Based on StarSpace + previous bidding information

Semi-Automated Bidding: Results

- Quality of results improved over time
 - Prior bidding info -> Significant improvement
- Never reached the point of being fully trustworthy
- OTOH, significant fraction of PC didn't enter bidding preferences
- Example feedback in December:
 - *"FYI, I continue to be less than thrilled."*
 - *"It seems that automatic preference generation has been doing worse than the topic score in the last couple of rounds. Did you use a homepage of a machine learning expert to generate my preferences?"*
 - *"The automated system is good at removing most of the things I do not want to review... It just regularly puts things I know nothing about on top of my list."*
 - *"Auto-bidding is getting better! Mostly, the top-rated papers for me were ones I could review."*
 - *"The inference of preferences is actually pretty good :-)"*
 - *"The automated bids look good to me."*



Other Logistical Challenges

- HotCRP
 - Not designed for monthly deadlines
 - Web chair Ben Andow helped twist it to our purposes
 - Relied heavily on tags
 - Mostly worked, but not ideal – more automation needed
 - No support for automated bidding
 - Accomplished via SQL hacking
 - Integrated solution with more controls exposed to PC would be better
- Deadline timing
 - Review crunch for September round coincided with NSF deadlines
 - Review crunch for November submissions bumped into December holidays
- Keeping reviewers on track & aware of deadlines
- Continual chairing

Anecdotal Evidence: Mostly positive

- *“Reviews for the paper were amazing, some of the best I’ve ever seen. The suggestions for the revision were exactly right and helped motivate the student involved to push the work to completion.”*
- Paper wasn’t quite ready so waited a month
- Two students submitted in November and two in December
- Triaging papers sent to other conferences

Submission Statistics Summary

- Total submissions: 549
- Well-formed submissions: 533
- Eventually accepted: 66-68 (14% acceptance rate)
- Accepted in time for 2018: 63 (12% acceptance rate)

Comparison With Previous Years

Submissions

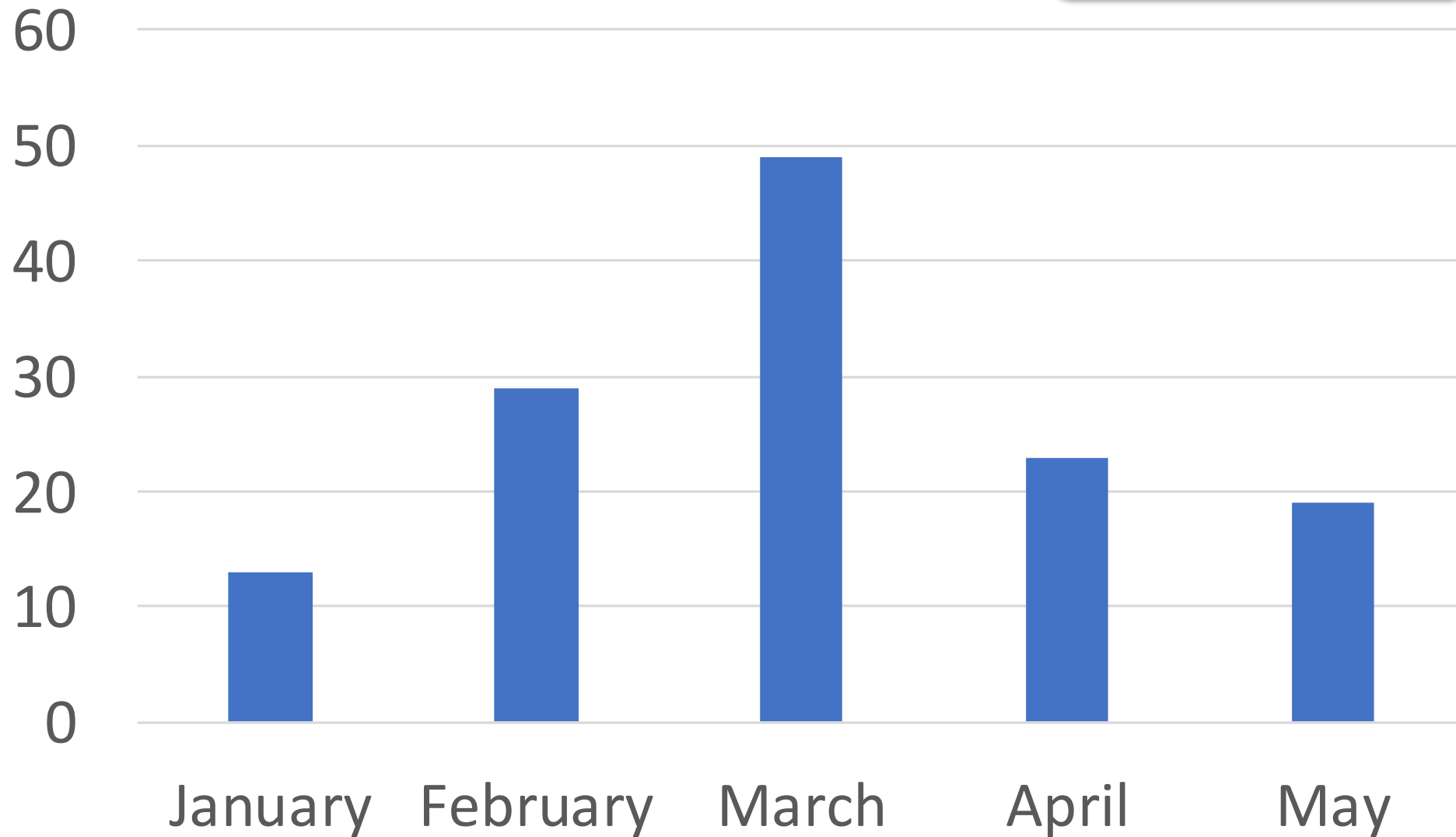
+27%

Acceptance rate

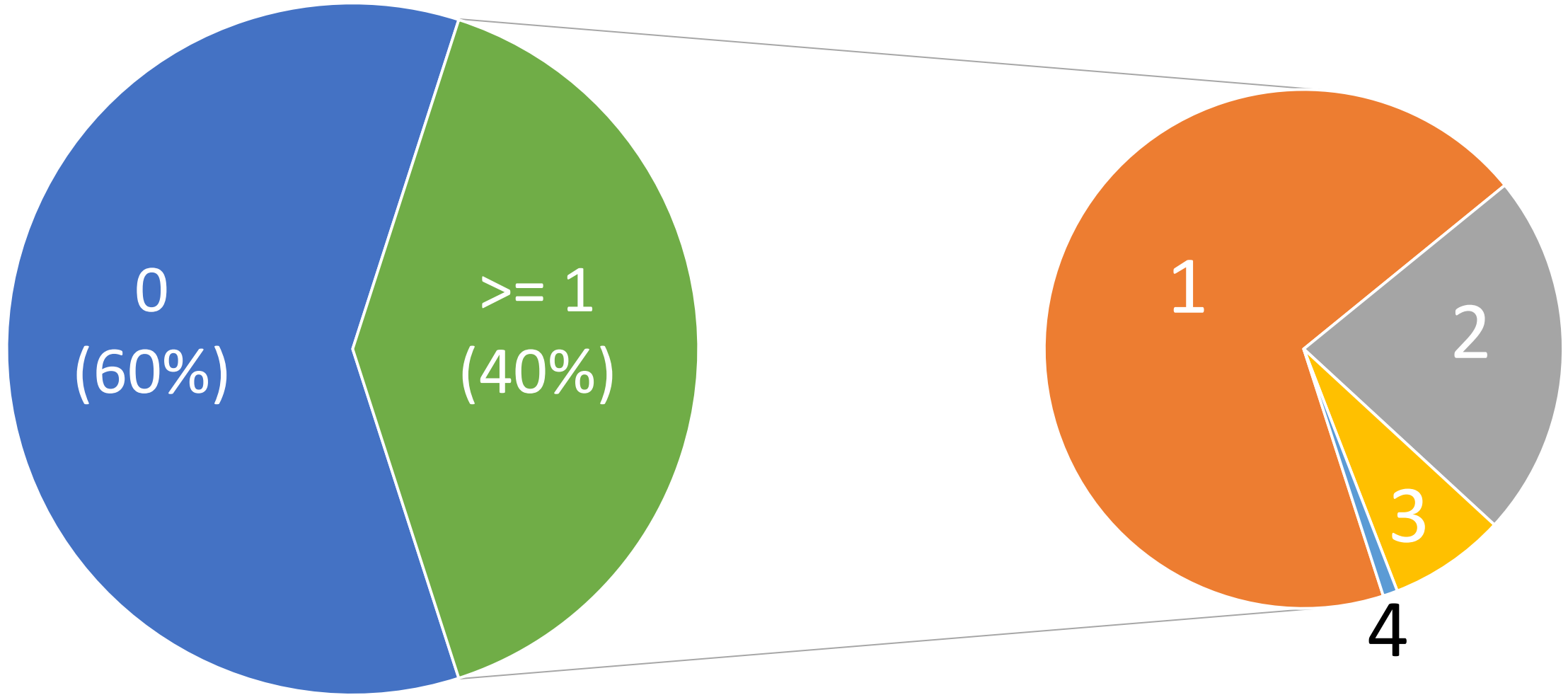


2019 Submissions

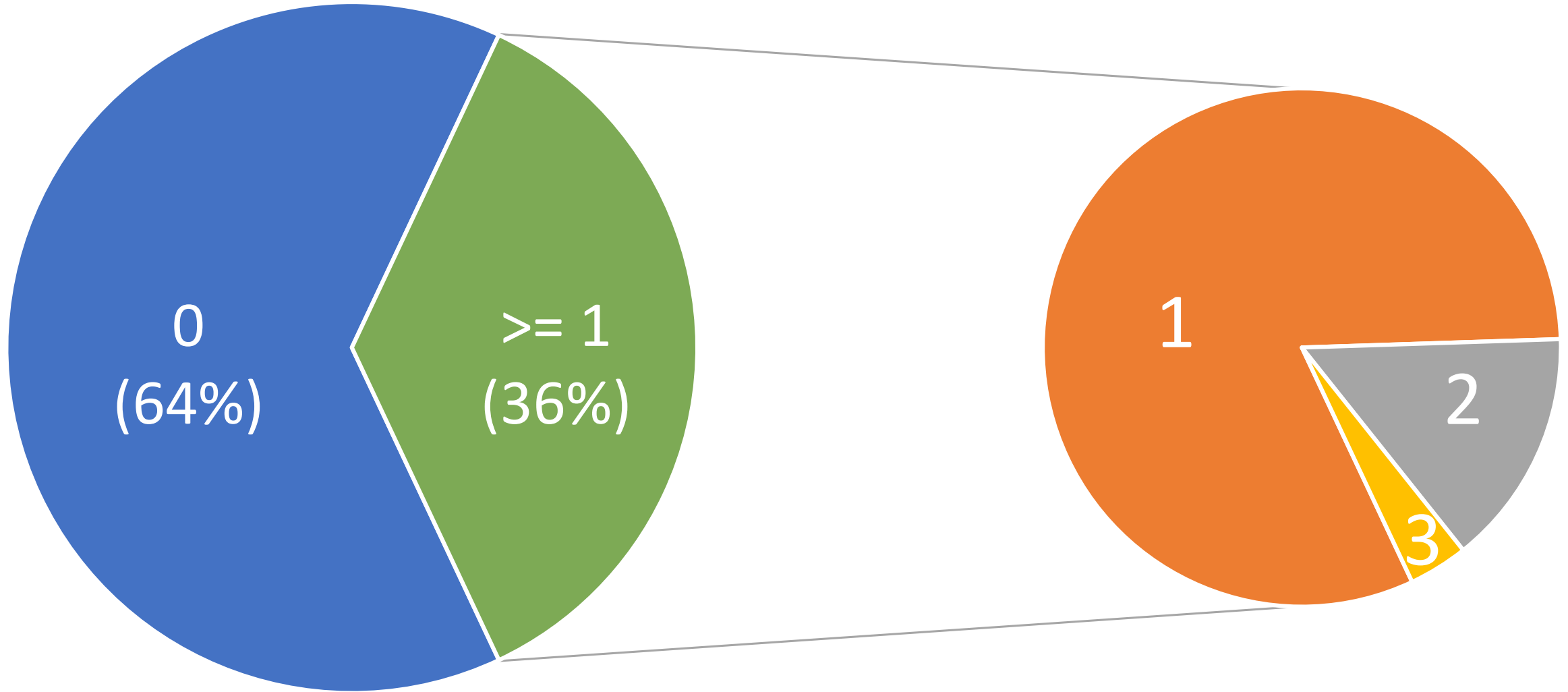
Total: 133



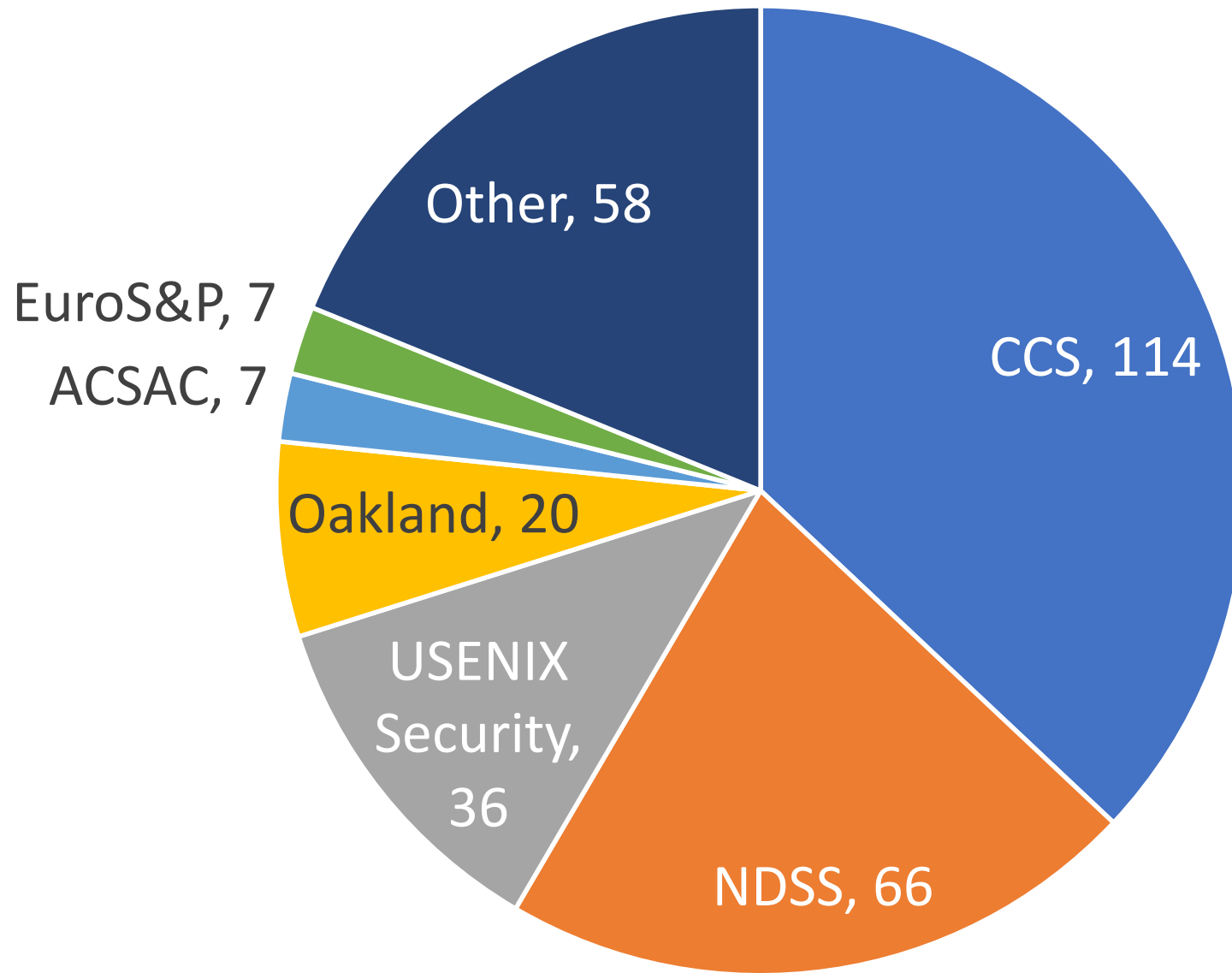
Number of Previous Submissions in Last Year: All



Number of Previous Submissions in Last Year: Pos



Previous Submission Venues



Impact on Other Venues?

- Euro S&P 2018 received 47 fewer submissions
 - 145 vs 192
- NDSS 2018 received 73 fewer submissions
 - 350 vs 423
- USENIX Security 2018 received 48 fewer submissions
 - 524 vs 572
- CCS 2018 received 33 fewer submissions
 - 809 vs 842

Revisions: Risks

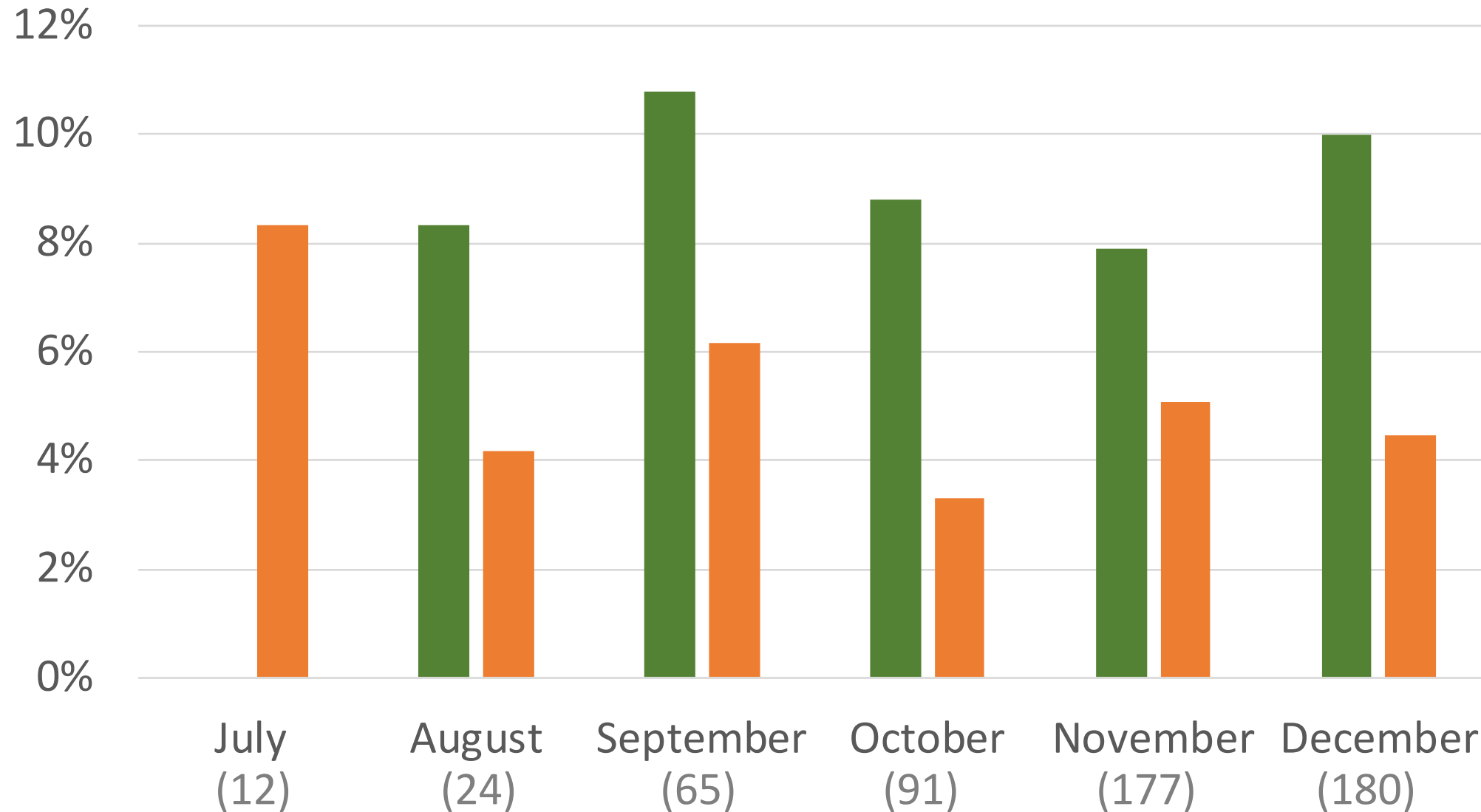
- Used too conservatively: Little change in review process
- Used too liberally
 - Good papers:
 - Take longer to publish than in old model (journal fail)
 - Bad papers:
 - Reviewers spend more time on a paper that needs more than 3-months of work
 - Authors get their hopes up and resubmit rather than trying new venue
 - Reviewers feel pressured to accept mediocre papers (journal fail)
 - Evidence would be nearly all revisions accepted

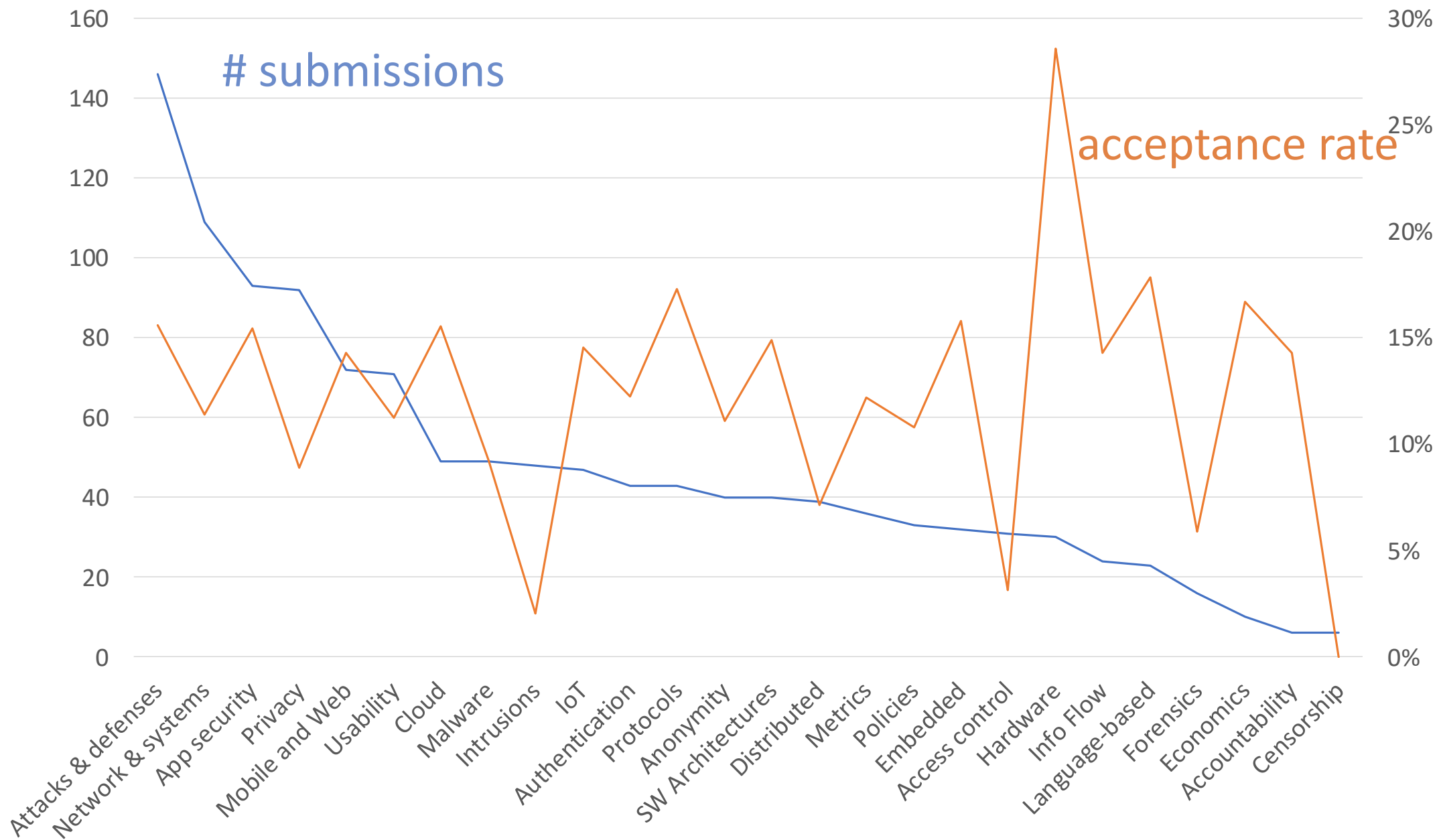
Revisions: Guidelines to PC

- Use Revise sparingly:
 - If a paper would have been accepted previously it should still be an Accept
 - If nobody likes the paper, it should be a Reject
 - If it's an exciting/good papers except for X, where X is a missing experiment, a crucial application then Revise is an option
- Revise shouldn't be a fishing expedition to maybe possibly find a good paper by having brilliant new insights
- Need a clear list of conditions under which paper would be acceptable S&P paper
 - If authors meet conditions, should be ready/excited to accept
 - We do not want multiple rounds of revision

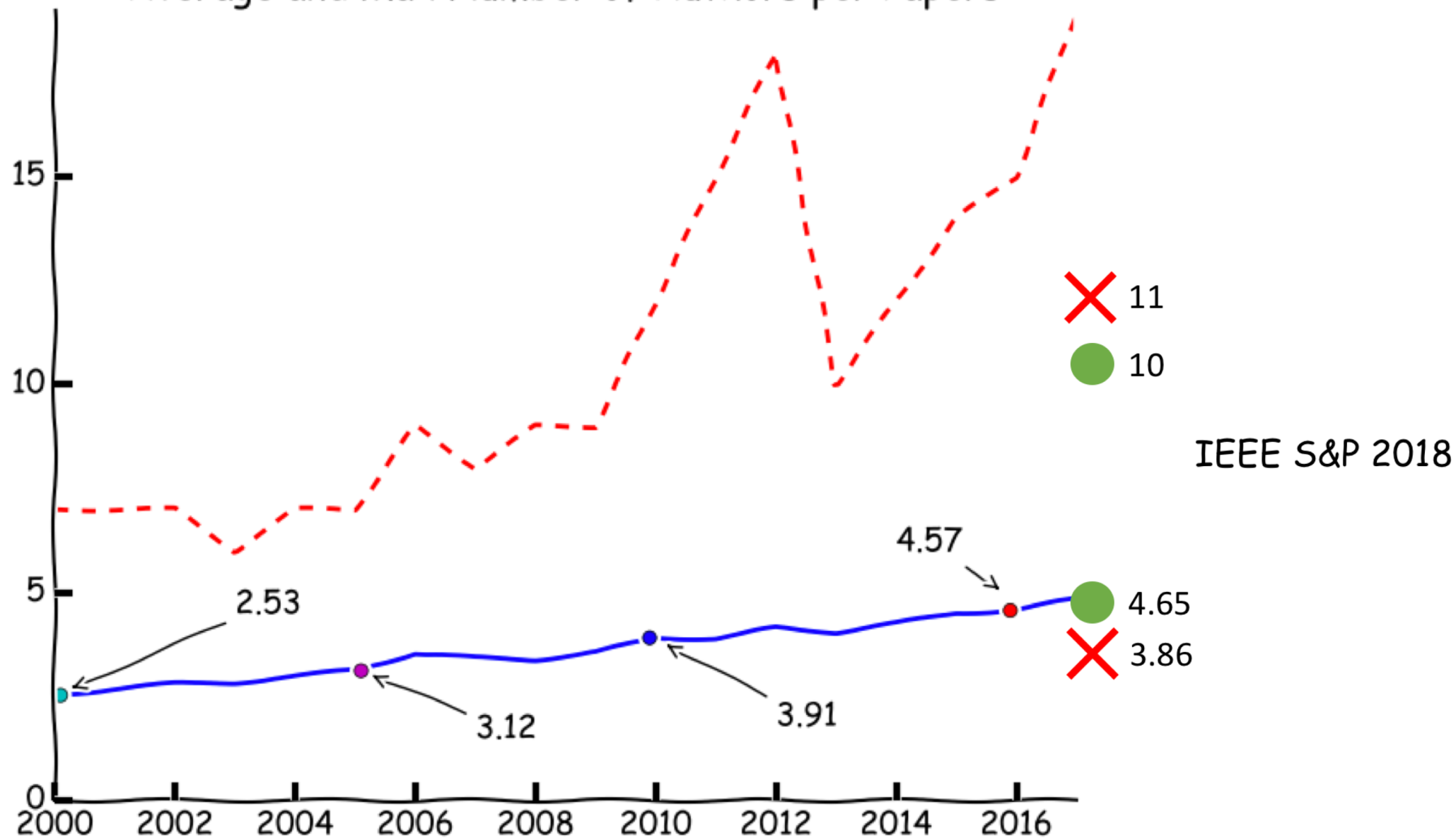
Decisions on New Submissions

■ Accept Rate ■ Revise Rate

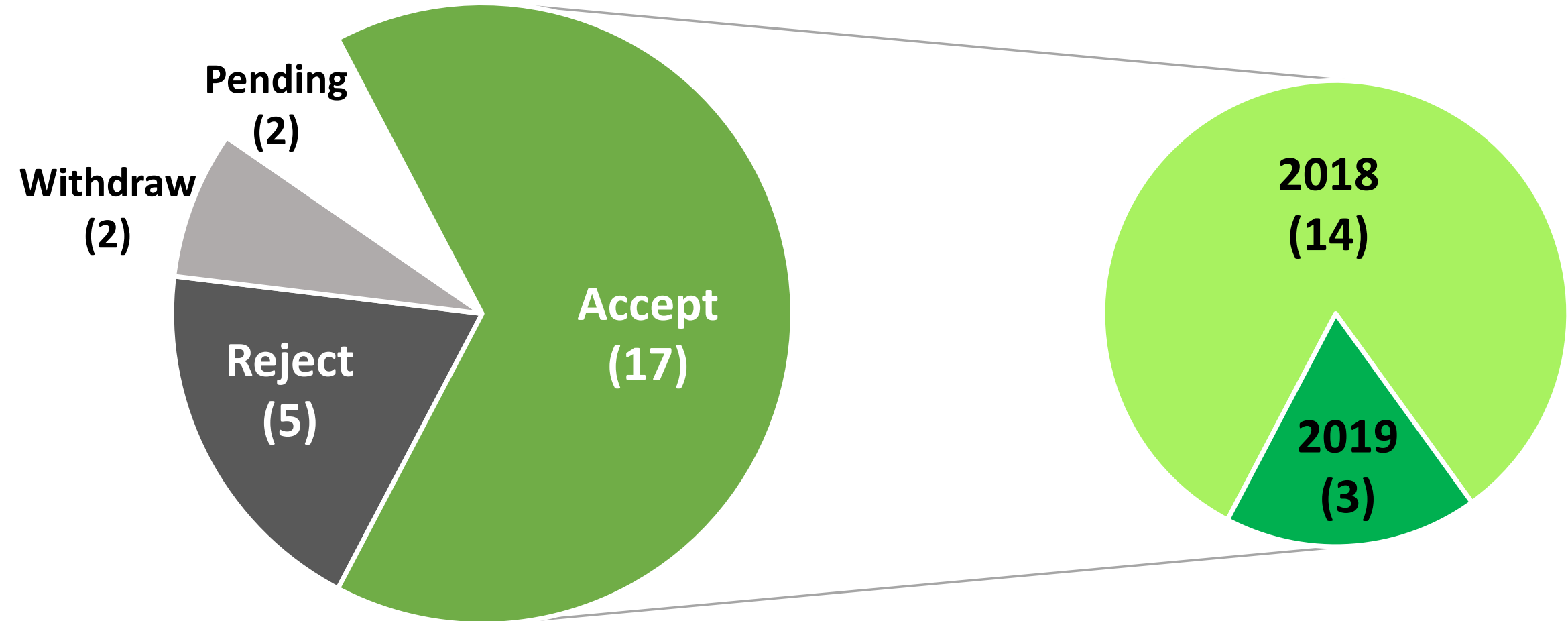




Average and Max Number of Authors per Papers



Decisions on Revisions



Revisions: Reflections

- Revisions != Shepherding
 - Many papers that the PC said absolutely could not be accepted on first submission, but that turned into very nice revisions
 - Most required additional experiments, or major changes
 - 65% revisions accepted vs ~100% for shepherding
- Several papers PC initially suggested Revise, but with no clear list of expectations
 - Result: Reject
- Overall acceptance rate comparable to previous years
 - => Revised too many good papers?
 - => Revised too few borderline papers?

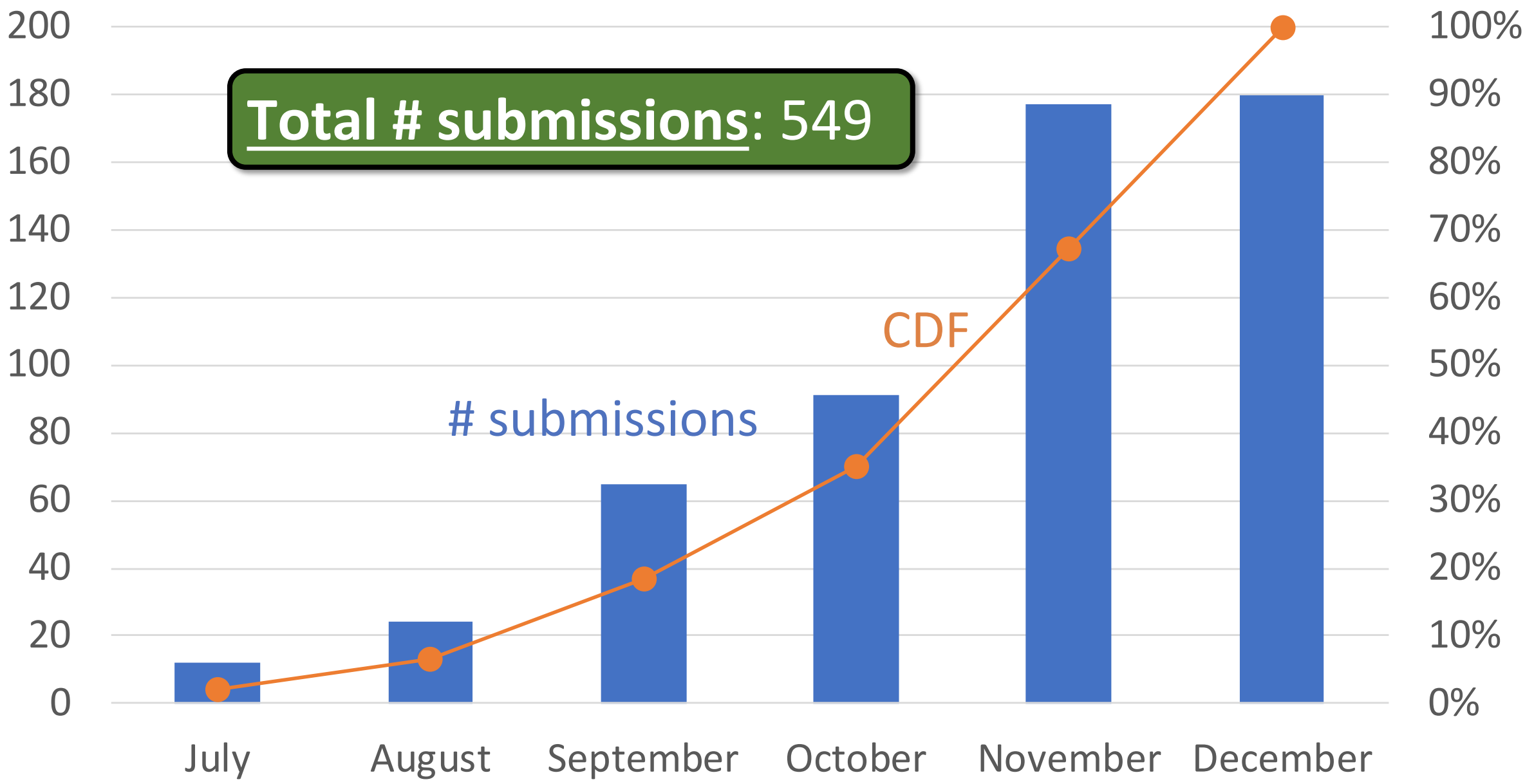
Review Process

- Increased PC size by ~50% (60 -> 87)
 - # of women doubled (went from 12% to 16% of the PC)
- Split PC into two pools with comparable skill sets
- Only assigned new submissions to one pool in a given month
 - Most reviewers got new papers ***every other month***
- Timeliness
 - All 549 decisions returned within 2 months of new submission
 - All but one revise decisions within 1 month of submission
- Review load
 - Significantly lighter
 - Median of 18 and max of 21 (vs avg of 24 in 2017)
 - Less spread out than hoped

Submissions

CDF

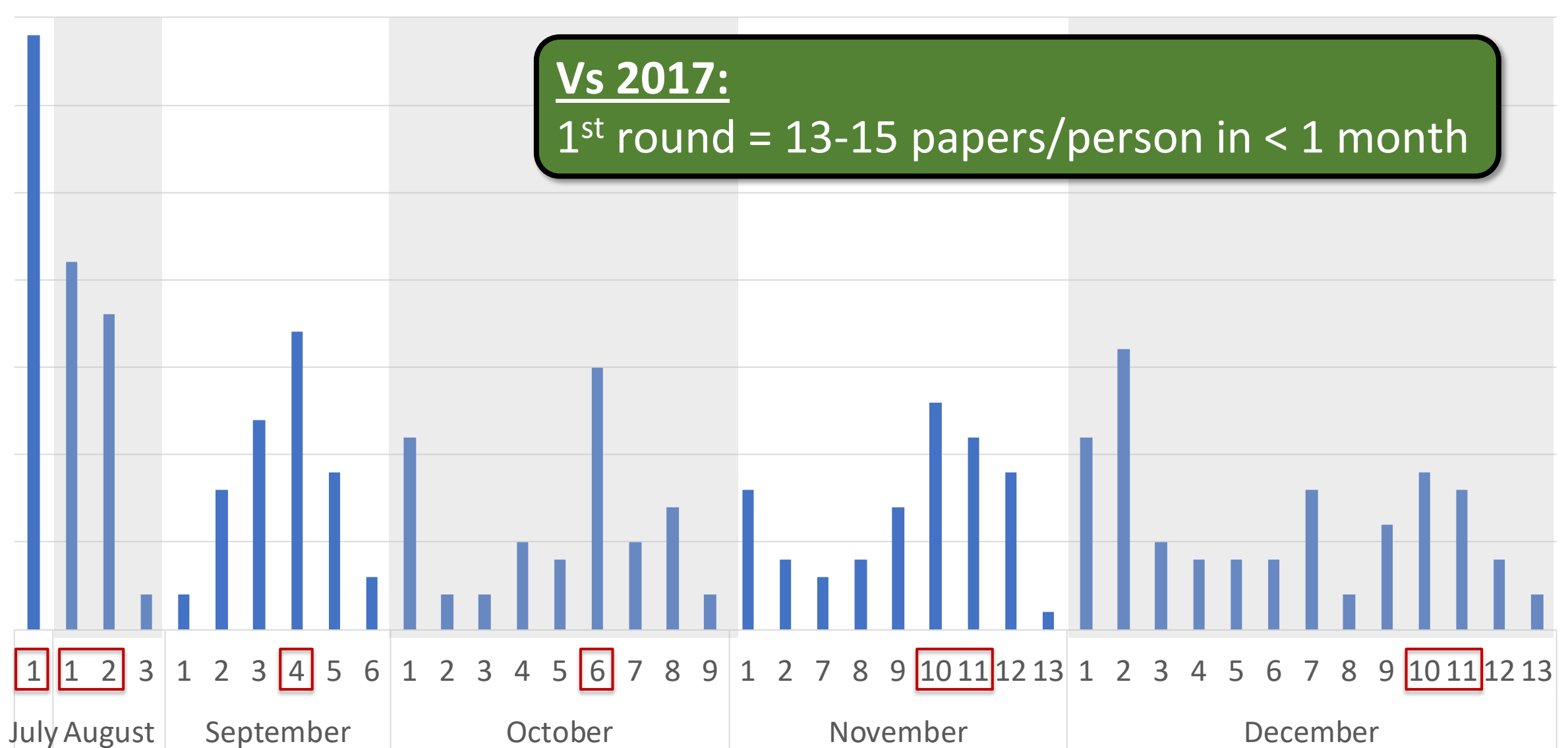
Total # submissions: 549



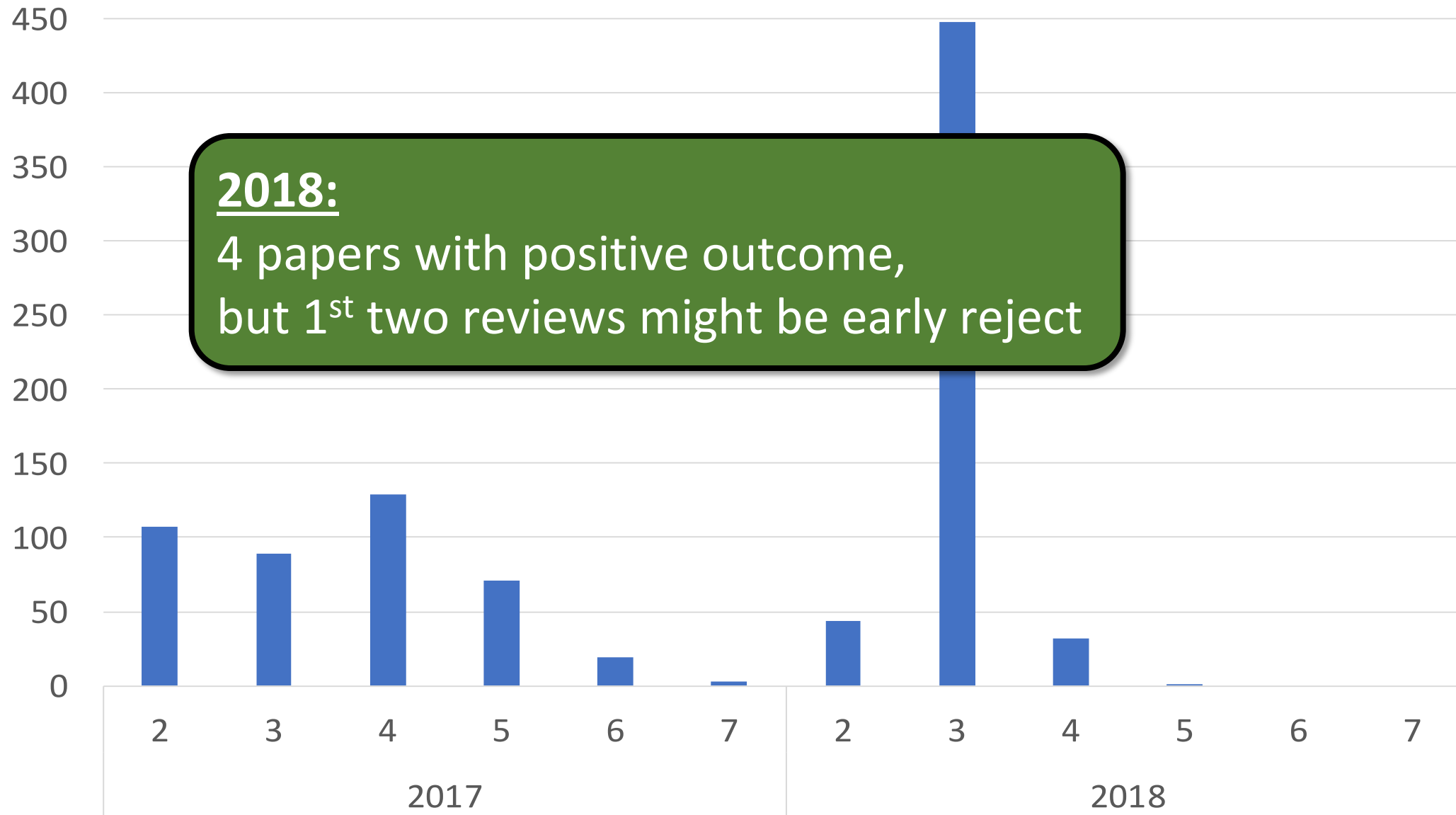
Assignment counts / month

Vs 2017:

1st round = 13-15 papers/person in < 1 month



Reviews/Paper



Reviewer Pessimism

- Top overall merit score: “6: One of the best in my pile”
 - Number of PC members using this score: 9 (out of 87)
- Number of papers with at least one 5 or higher: 54 (out of 547)
- Average of average scores: 4.03 (std. dev. 0.43)

PC Survey Results

Special thanks to Lujo Bauer!

- Same questions as in 2017 survey
- 44 complete responses = 51% response rate (vs 77% in 2017)
- These are preliminary results
 - No correction for reviewing load, scores assigned, etc.
 - For comparisons with 2017, limited significance analysis

Do you feel confident that the review process led to the top papers being selected for publication (i.e., overall, not just the ones you reviewed)?

- 70% positive (vs 85%)
- 16% negative (vs 6%)
- 14% unsure (vs 8%)

Not significantly different

From your perspective as a reviewer, was the page limit on papers too generous, about right, or too stingy?

- 77% About right (vs. 54%)
- 20% Too generous (vs. 35%)
- 2% Too stingy (vs. 0%)
- 0%. Much too generous (vs 10%)

How much time (in hours) did you spend, on average, on each paper you reviewed (including reading the paper, writing the review, reading and responding to reviews online, and in-person discussion)?

- Min: 1
- Max: 8
- Avg: 3.9
- Median: 4 (same as 2017)

How much do you think the **(virtual) PC meetings** contributed to **the quality** of the program (compared to just online discussion)?

- Improved: 59% (vs. 72%)
- Hurt: 9% (vs. 2%)

Not significantly different

For the 2018 symposium, we had monthly review cycles and we tried to assign papers to a given reviewer only every other cycle. Given that experience, would you

- Keep things the same: 50%
- Fewer cycles with more papers: 27%
- More cycles with fewer papers: 20%

For the 2018 symposium, we introduced the option to decide a paper should be revised, rather than accepted or rejected. Did you feel you had a good understanding of the criteria for a paper to receive a revise decision?

- Yes (75%)
- No (25%)

Author Survey Results

Special thanks to Lujo Bauer!

- Same questions as in 2017 survey
- 275 responses from 1115 authors = 25% response rate (22% '17)
- These are preliminary results
 - No correction for paper decision, submission date, etc.
 - For comparisons with 2017, limited significance analysis

Would the paper you submitted have been substantially stronger if you could have spent additional time on it?

- Yes: 39% (vs. 46%)
- No: 51% (vs. 45%)

Not significantly different

The reviews I received mostly evaluated my work objectively. I.e., the decision seemed to be based mostly on facts and not mostly on reviewers' subjective opinions (e.g., disliking an area or style of research).

- Agree: 56% (vs. 60%)
- Disagree: 33% (vs. 27%)

Not significantly different

The reviews I received were constructive and respectful (even if I disagreed with the subjective assessments or final outcome).

- Agree: 65% (vs. 77%)
- Disagree: 17% (vs. 14%)

Not significantly different

Compared with previous years of this conference, the overall quality of the reviews I received were:

- No previous submission: 44% (vs. 44%)
- Better: 12% (vs. 16%)
- About the same: 25% (vs. 31%)
- Worse: 17% (vs. 14%)

Not significantly different

Revisiting the Original Motivation

– Permit submission and acceptance of results any time they're ready



- Frequent submission opportunities



- Rapid turnaround to published paper



- Reduce incentives to “cram” before deadlines and submit half-baked work

– Improve quality of papers and review process



- Start with papers written under less time pressure



- Facilitate a more collaborative process between reviewers and authors

- Smooth the reviewing workload to a low-volume, ongoing effort

Conclusions

- Still less than 1 year into the new system
- Mostly positive impressions
- More tweaks possible
 - More reviews for top papers?
 - More liberal Revise policy?
 - Fewer deadlines?
 - More incentives to submit before November/December?
- Still a work in progress!

Thank you!
parno@cmu.edu