

You Must Be Joking...

The most serious column strikes back

Michalis Faloutsos, UCR
featuring Kevin Almeroth, UCSB

Despite a few emails of constructive criticism (mainly death threats), a deportation scare, and a paternity lawsuit, the first column got the response any first column hopes for: it went largely unnoticed. A triumph. So, on we go.

First, some good news. Five SIGCOMM members have been selected as ACM Fellows: Tom Anderson, Christophe Diot, T.V. Lakshman, Walter Willinger, and Hui Zhang. The awards will be presented at the annual ACM Awards Banquet in San Francisco on Saturday, May 20, 2006. Among them, I would like to distinguish Christophe Diot, for the sole reason that he is the CCR Editor, and therefore, the boss as far as this editorial column is concerned. “Bravo, monsieur le directeur.” All jokes apart, warm congratulations to all the new ACM Fellows. I am sure you all recognize their names. If not, you need to stop reading columns like this one and read more papers.

A social commentary (also known as bitching and complaining) follows. I love this job.

1. A CLEAN SLATE APPROACH TO REVIEWING PAPERS

The system for peer reviewing academic papers is not scalable, and things are not getting any better. If it is any consolation, other disciplines seems to have worse problems. The recent scandals in the stem cell research have brought forward many questions. One question that received less attention is whether the peer-reviewers did their work. We extrapolate from that: what exactly is the responsibility of the reviewers? Maybe the better question to ask is whether it is really the reviewers’ job to identify intentionally falsified or fabricated data? This is a question that we do not attempt to answer here, but we included it to add drama and depth to this article.

The main point is that reviewing is barely working now, and it will likely collapse soon: the system is not scalable. And this happens to a community which has “scalability” as the reigning king of design. To make things worse, the system is buggy and brittle, and it lends itself to abuse. The decreasing acceptance rates create a death spiral: bitter stressed out people with less time to devote to reviews will only make things worse.

Our question is: can we review the papers effectively, if we don't really review the reviews?

Reviewers are responsible competent colleagues who fully understand the responsibility they undertake, and they are willing to spend as much time is needed to do thorough reviews. Clearly, each one of us falls in this category. It is the reviewers of our papers that are incompetent, biased, lazy clowns, right? So it is not me or you, it is the others. This paradox would make Zenon cry from joy (if you don't know Zenon's paradox of Achilles racing the turtle, avoid betting in horse-races).

Currently, the only mechanism for judging the quality of reviews and reviewers is through the Program Committee (PC) meeting. The quality and accuracy of reviews are vetted and the qualifications of the reviewer are at least examined. Of course, as the number of workshops and conferences has exploded only a small percentage still continue to have PC meetings. And, of course, let's not talk about the attendance level at these events or what *really* happens when deciding what papers should be accepted. How far are we really from a system where reviewers just write reviews and the program chairs' only job is to try and normalize quantitative scores?

Don't get me started on journal reviews. We wait for a year to get a review that often reads: good idea, but they need to do more simulations, therefore, reject. Let alone that in the duration of a year, the student has graduated to make double our salary in industry. Sounds familiar?

The current reviewing system has many structural inadequacies (which is a convoluted way of saying it sucks). Here is the core of the problem:

1. *Minimal quality checking capability*: there is no real good mechanism to detect and discard bad reviews. Sure, you can detect a one-line review, and potentially disregard it (do we, really?). However, what do we do with a 3-paragraph review which boils down to the vague statement: “I don't like it” without pointing to flaws or providing suggestions to improve the paper (what a foreign thought that is)?

2. *Blind reviewing: is it really blind?* I am sure we all have our stories here. The technical report that is publicly available with the same title, the friend of the author who does not have a conflict so s/he stays in the discussion, or just the nature of the work and the writing style.

3. *Impunity: Reviewers are untouchable.* They don't get called on their bad reviews, and there is not even much memory in the system.

4. *Reviewer collusion and intentional misbehavior.* Not very wide spread admittedly, but non-negligible either. How big is this problem? Here is a measurement paper we would like to see.

5. *Lack of incentives to do a good review.* Sure, we all want to be on the prestigious TPCs, but there is no reward for doing a good review (or a timely review, when it comes to journals). Can we get "reviewing air-miles" somehow?

Well certainly, a few experimental ideas have been tried. For example, various conferences have experimented with ideas like online reviewer discussion, discussion summaries, author rebuttals, multiple reviewing rounds, double-blind submissions, limited submissions, and no-PC-member submission. Has this worked? Don't you think we would know if it had?

If we envision the possible set of alternatives for the reviewing process as a spectrum, Computer Science has largely been navigating with the principle: hiding information improves things. It sounds reasonable until we realize that you it is very hard to ensure information leaks as we mentioned above.

As a thought experiment, let's consider then the opposite end of the spectrum: transparent reviewing. You know who reviews your papers, and you can read their reviews on-line. Part of the rationale is that ensuring blind review is difficult to enforce, ensuring that we all know is enforceable.

A shocker, no?

At first it sounds stupid, because we are conditioned to think otherwise. But let's not rush to conclusions. How much worse can it really get? Just joking. Partially. The point is that we have a problem currently, and it may be worth of thinking a really disruptive solution.

System Design. This is how we envision it will work. All reviews and the names of the reviewers will be made public. Then, we will archive all the reviews for life and have tools to ask interesting queries and do correlations and reviewer profiles.

Scary, eh? It will definitely make you think twice before writing: "Obviously the authors do not have a clue about networking." Here are some other interesting properties of this approach.

Review quality: This is going to be the biggest advantage: the reviewing will improve by leaps and bounds both in quality and also in attitude. Reviewers will be forced to be technically accurate, specific, polite, and

more inclined to provide useful feedback, which what reviews should be.

Accountability: Reviewers will be forced to take ownership of their reviews: they can't hide behind the anonymity.

Collusion prevention: It will curb reviewing-related crime. Reviews stay and they form your "credit report". Whether you are a serial paper-killer, or a biased reviewer for or against particular people, it will show.

Transparency: it takes away the guessing game of who did what, and whether a PC member will squeal and hint at who possible may have fought against that paper. Everybody knows, so at least this levels the field.

Normalization: This can be used as a way to normalize the review scores. If nothing else, it can be an indication of the strictness of a reviewer, which the PC chairs can use at their discretion or as tie breaker.

The first obvious drawback is the vengeful behavior of people. What makes things worse is that some people have more important positions more often than others. If you cross them, they may haunt you for life. And, yes, forget that reference letter too. Can we assume that these bad people will be very few, and most of us will just see the importance of being impartial? Don't make me laugh.

Our thinking at this point is that there may be ways to get around this. First, we could have a correction coefficient for powerful people to correct the bias of fear. Second, instead of saying accept reject, we could just order the allocated papers we get. This way the reviewer is not saying that your paper is bad, s/he is just saying that s/he thought 5 papers were more interesting, but s/he still wish dearly that your paper was also accepted.

The second potential problem: people who will try to "flatter" key people. The solution here is the "credit report", the shame, and the peer pressure.

Here is another idea: a reviewers credit rating could also be tied to how well the papers you championed do later. It can vary from accuracy, bias (many different categories i.e. per topic, by author prestige, by author nationality etc), timeliness. Then, this score could be used as an extra property of your professional standing. For example, one would consult the credit for picking reviewers.

Will this approach solve all problems? Certainly not. However, the right question to ask is whether it will work better from what we have now.

In conclusion, as community, we are trying to revolutionize wireless communications, rethink the networking architecture, think outside the box etc. We could use some of this revolutionary momentum to solve this other immediate problem. Let's think about it a bit.