letter to the editor

To Rob Kolstad:

We've long admired Rob Kolstad. He is a down-to-earth guy who is not afraid to say what's on his mind. And what's on his mind is invariably worth hearing, unless you happen to be the poor soul at the podium when he's stating his mind at the aisle microphone. However, when we read his editorial in last month's *;login:*, the one where he comes out in favor of ISPs blocking port 25 of zombie machines being used to send spam, we decided we needed to do something we thought we would never do: suggest that Rob Kolstad is wrong.

Well, maybe "wrong" is overstating it. It's not that we think the selective type of blocking Rob is advocating won't throttle the zombies the way it's supposed to do, and it's not that we think that adding this to the anti-spammers' bag of tricks is necessarily a bad thing. But neither is it true that this type of blocking or even the bag-of-tricks approach as a whole is the most effective or efficient way to attack the problem, and Rob should know that.

He should know it because, just five months ago, he sat in the audience at LISA '04 when we were presented with the Best Paper Award for our work on spam filtering. Given that we won the award, you might jump to the conclusion that we created some genius filter, or that our work was highly theoretical or very complex. Given that multiple people at the conference were heard to remark that we had "solved the problem," you might think that we'd created some magical solution, and it just hasn't reached the practical masses yet.

You might think that way, but the opposite is true. We didn't create anything. We won the award with an implementation paper. That is, we did what sysadmins are supposed to do: we read the prior work, took some commonly available components, made the adjustments necessary for our own environment, and implemented this cobbled-together solution. Our solution was to implement a simple Bayesian filter (an obsolete one, by today's standards), and—surprise!—it worked. Our paper just documents our implementation methodology and describes how you can use it at your company *right now* to effectively solve the spam problem for your user base.

Yet here is Rob Kolstad expressing his frustration at spam (a frustration we all share) and recommending, not a systematic or comprehensive solution, but yet another filter-of-the-day approach to today's most popular spammer trick, sending via zombies. Now, Rob certainly isn't alone in suggesting that we need to look to solutions other than filtering, including extreme measures such as port 25 blocks, selective and otherwise. There are a veritable circus of people like Rob—people much smarter than we are—screaming at the top of their lungs about how we can't win this fight with the "outdated" protocols we have today. They assure us that SMTP needs to be rewritten, email needs to be charged for, and access to the Internet must be censored. So how is it possible that two sysadmins at a healthcare co-op have already functionally solved this problem for their user base with a two-year-old version of a Bayesian filter? How is it possible, given that LISA '04 also had an invited Ph.D. from Microsoft claiming that "the problem with email is that it's free"?

The answer is that today's sysadmins appear to have acquired something akin to Attention Deficit Disorder wrapped in a "Somebody Else's Problem" field when it comes to the spam problem. We see the symptoms in pre-

sentations and conversations at the conferences we attend, in the papers we read, and in articles like Rob's. We hear and see sysadmins discussing federal anti-spam case law. Sysadmins demanding SMTP protocol rewrites and IETF draft acceptances. Sysadmins begging ISPs to shut off core Internet functionality for their users. Sysadmins talking about everyone in the world needing to adopt DNS hacks to send email. In short, a whole lot of sysadmins demanding that other people solve the problem for them and just tell them what to do, and a whole lot of "flavor of the week" and "reinvent everything" approaches born out of these demands. We *want* them to tell us that it isn't our problem and to explain how it's the current way things work that's broken. And if they want to charge us for some black box or for recommending that we redesign (or even turn off) core functionality, that's fine too.

What happened here? We are system administrators. It is our job to solve these kinds of problems using the tools we have available, without breaking interoperability. It's our job, and we used to enjoy it. We used to be good at it, too.

If there was a single, non-utilitarian point in our paper, a moral to our 20-page ramble, it was that we don't need censorship, FCC regulations, protocol changes, protocol kludges such as SPF, or ever-smarter learning algorithms to solve this problem. What we need is more system administrators doing competent implementations of good learning filters such as Bogofilter, crm114, and DSPAM.

We know that filters have been "beaten." There's wordlist poisoning, nefarious HTML tricks, microspam, an arms race, etc. etc. We've heard the professors and fellows, and professional smart people. We know it's impossible. The difference between them and us is that we see these attacks daily, we see them in the wild, and we've seen them in real time for two years now. And we'll summarize that experience in three words: They don't work. Despite so many people assuming these filters have been broken, the data just isn't there. To our knowledge there is not a single published paper that empirically demonstrates fatal flaws in these filters or shows them to be less than adequate at solving the problem when they are implemented using a sound methodology. Quite the contrary. Yes, they can be made to have sub-par performance when people are lazy and try to cut corners, but if the cost of solving the problem is a bit of homework and elbow grease, system administrators are the last ones who should be complaining.

Yes, we know that filters aren't the utopian solution because the spam isn't blocked at the sender. But while we argue about the ideal way to do that, users are getting ever more inundated with spam that is increasingly offensive in nature, and many of them are abandoning email altogether. We have something that may well work permanently to reduce spammers' ability to harass us and put us back on the offensive. We should take full advantage of this first and then worry about cutting spam off at the source.

Please, let's turn down the volume knob a few notches; let's help as many sysadmins as we can to get good filtering implementations; let's do our jobs. Let's see how deep the filtering rabbit hole goes for real before we chase after the next shiny idea that offers to solve the problem for us, only at the cost of a little bit of core functionality.

DAVID JOSEPHSEN AND JEREMY BLOSSER

dave@homer.cymry.org

Rob Kolstad replies:

I stand by my comment that ISPs should take measures to stop spam at its source. Filtering at the delivery point does protect end users—potentially at a high level—but it still incurs what I believe are unacceptable costs all down the line: bandwidth, CPU cycles, administrative personnel costs, and (worst of all) diluting the effectiveness of email as a powerful tool. Email's effectiveness is diluted by losing important messages (false positive detection) and wasting the time of readers (letting spam through). Of course, filtering only works for those who have filters installed. The rest continue to suffer the scourge, a scourge promulgated for no ethical reason that I can discern.

I have good results with a Bayesian filtering solution on my system, but I think that's only the beginning of a total solution.