

Taking Another Look at the Spam Problem

by *John C. Klensin*

Reprinted from The Internet Protocol Journal (IPJ), Volume 8, No. 4, December, 2005
IPJ is a quarterly technical journal published by Cisco Systems. See www.cisco.com/ipj

The problem of unsolicited bulk e-mail on the Internet has been widely discussed, and many classes of solutions have been proposed. Dave Crocker's article discusses some of the background for the solutions generally, points to a semi-humorous list of ways in which proposed approaches fail, and compares several approaches based on source authentication. This article takes a somewhat contrarian view. It argues specifically that the traditional models for defining technological solutions and then letting the policy and legal communities work out the details of how to utilize them are seriously wrong in this particular case and that partially-effective methods of fighting spam actually cause more spam.

This article makes two main suggestions. First, attempts to design technological countermeasures to spam without a clear understanding of how far, and in what directions, the setters of social policy are willing to go are futile. The requirement is not just that there be social recognition that a problem exists. In order to design effective technological countermeasures with predictable and acceptable side-effects, we must first understand what measures society is willing to take—what laws it is willing to pass and enforce to make spam a criminal or civilly-punishable act—to set an appropriate context and set of boundary conditions. Without those conditions, design of technological countermeasures is likely to constitute poor engineering practice, not just futility. Second, deployment of spam counter-measures that are not completely effective largely shifts the burdens of spam from one recipient population to another while *increasing* the total amount of spam on the network.

His analysis and mine agree on several critical points. Solutions that discard important characteristics of today's e-mail environment permanently in order to make some short-term gains against spam are not acceptable. Approaches that require drastic and simultaneous changes to the ways in which e-mail works in order to function are not going anywhere. There is a difference between legitimate businesses who have decided, within the limits of existing legislation, to engage in mass, unsolicited, electronic mailings to promote their products and those bulk mailers who prefer to cover their tracks, hide linkages between sending addresses, hosts, and web sites (or create deceptive ones), and who use zombie mailers and other ways to avoid cost and detection. We also agree that spammers, or their tool suppliers, are creative, technically knowledgeable, and able to react much more quickly than the spamfighting community (especially the standards-based part of that community) to changes in operating conditions and countermeasures.

I suggest a further guideline to help us think about the problem: however small they might be on a per-message basis, there are costs associated with sending e-mail and costs associated with receiving it and eliminating undesirable content.

If an anti-spam "solution" is developed that permits the spammers to vastly increase the costs to the recipients without a proportionate increase in their own costs, that

solution is not tenable. A serious effort to predict the impact of a proposed solution to spam, including costs to the end user and load on the network as the spammers adapt to it, should be a critical component of such efforts. But, while equivalent analyses of measures, likely responses, and countermeasures are standard with any (other) technique designed to enhance network security, they have been largely absent when new technological approaches to spam are proposed.

This is a different aspect of the so-called “arms race” problem. In a classic arms race, no one can really win, as Dave points out. But, more important, when such races stop, it is only because one party simply stops, is forced out of the game by external pressures, or becomes exhausted economically. As long as there are no economic constraints, every escalation is met with a counter-escalation, which is met with a counter-counter-escalation, and so on. It is this positive feedback cycle that characterizes a true arms race. The battle against spam demonstrates a particularly unfortunate variation on that pattern in which the incremental economic costs of trying to deploy new spam abatement measures appear to be much more severe than the costs to the spammers of the most obvious counter-measure to improved spam abatement procedures, simply sending out more traffic. This is discussed further and in context below.

Social Problems and Technological Solutions

In the technical and protocol design community, our normal model is to develop technology and then use it to inform the policy, social, and legal parts of the society who then need to sort things out on their side. One of the classic arguments for this approach, which does not seem relevant to the spam situation, is that the potential use or misuse of a technology will not, and should not, constrain its development. For spam, the situation appears to be exactly reversed: we need to understand what is feasible and plausible from social, political, legal, and regulatory standpoints in order to define the engineering solution space. If we do not know what behaviors society is willing to make illegal or subject to effective civil action and whether it is willing to enforce those laws or equivalent positions, we cannot adequately define the engineering solution space. That results, in turn, in a high risk of solving the wrong problem or an irrelevant one. Of course, recent history has shown a variety of irrelevant and costly solutions to spam proposed, and sometimes deployed.

The solution to spam is identical to the solution to most other significant social problems: society must determine that it is a problem, create effective rules prohibiting the problem, and then enforce those rules aggressively and consistently. Technical solutions that make it easier to identify spam and its sources can then be immensely useful, but they are only useful if designed to be effective within the framework set by those rules.

If, by contrast, societies are, in practice, unwilling to take effective social or legal action against spam and those who benefit from it, then this article suggests that anti-spam measures will tend to make the overall situation worse.

The question of spam beneficiaries provides a particularly good illustration of this point. So far, most legal systems in the world have taken the position that the act of

spamming is the offense (if there is any offense at all). Operating a domain or web site to which the spam recipient is directed to buy a product or obtain another benefit is rarely considered a problem by either law enforcement or by the relevant ISP. While establishing cause and effect—that the spam was authorized or encouraged by the web site owner—can be quite difficult, there has, appropriately, been little examination of tools to detect or identify beneficiaries because doing so seems pointless. On the other hand, on the same theory that it is more useful to try to arrest the drug importer than the street dealer, a different set of laws about beneficiaries and spam-authorizers—those who, in at least some cases, pay the spammers to spam—might dramatically change the landscape.

Reducing Spam by the Percentages

A new technique or group of techniques that claims to be beneficial can have either positive or negative value with regard to the amount of spam that gets through, either overall or to the mailbox or a particular sample user. A technique can also result in significant increases in the amount of network bandwidth or server resources consumed if it is neutral or better with regard to the end user mailbox. As long as the spammers can increase the number of messages they send out, almost arbitrarily and at low or zero marginal cost, the percentage of spam that is filtered out is ultimately irrelevant. The key measurement is how the amount of spam that gets through to some exemplar user (or a statistical aggregate of them) changes. That change pattern can be net either positive or negative. Suppose a technique is introduced that causes an initial small incremental reduction in the amount of spam delivered. The patterns of the last several years suggest that the spammers will respond by making a large increase in the amount of traffic they send out. Since the costs of doing so are very low, it would arguably be irrational for them to do anything else. If the increased volume is enough larger than the amount of spam the new technique was able to stop, there is a net loss to the Internet overall: the small improvement may represent a percentage decrease in the amount of spam that gets through, but the amount seen by the representative user increases and the percentage claims are largely irrelevant.

Unless whatever methods that are used in an attempt to reduce the amount of spam actually stop it at, or very near, the point of origin, the net effect on users is to shift the amount of spam received from those who have deployed the latest and most effective countermeasures to those who have not yet done so. The total amount of spam-related traffic on the network just continues to rise. And, since most countermeasures have costs—either in processing time or in software licensing fees—the cost burdens on end users also continue to rise.

This would seem to argue for methods that cut off spam traffic close to the source, but attempts to design such methods have been fairly unsuccessful, sometimes because of another policy problem: the spammers argue that some people like receiving unsolicited bulk commercial email so that cutting off bulk traffic near the point of origin prevents legitimate and desired traffic from transiting the network. Source-oriented techniques include not only technical approaches but efforts—by law or social pressure—to hold ISPs and mail providers responsible for all traffic emanating from their networks, thereby encouraging them to refuse to have spammers as customers, to aggressively enforce terms and conditions of service, and so on. The strongest advocates of the “blacklist” variation of those techniques continue to claim

that they are very effective although some others in the community are not completely convinced.

The House-Burglar Analogy

In the absence of a coordinated approach that is oriented toward legal or social enforcement, most anti-spam techniques appear to induce more spam on the network. They do this by making simply sending much more traffic out the most rational behavior for a spammer who is faced with an abatement technique to adopt. They may enable shifting the burden of dealing with that spam from one person to another—in the same way that aggressive locks and alarm systems on one house slightly increases the relative burglary risk to the less-protected neighbor—but, as Dave’s article points out, we have no realistic plan for making it too expensive for the spammers to simply increase output.

Deterrents to burglary work moderately well because they increase the costs (in time, sophistication of the required tools, and so on) to the burglar. Equally important, they increase the risks of being caught and punished. In the present spam environment in most countries, we have no effective mechanism to increase costs and, at least statistically, the odds of being effectively punished even if caught are insignificant.

Shifting Burdens and Creating Preferred Classes of E-mail

The argument Dave presents for authenticated mail is ultimately that it can get expedited handling while non-authenticated mail is put aside for other methods of spam detection. That approach could be immensely effective at expediting receipt of some mail by the recipients who apply the needed checks, at least until the spammers begin authenticating their mail in a way that tricks the trust-establishment techniques. Prioritization of some messages and content will be effective as long as the fraction of such messages remains relatively small relative to the total number of messages received. As the percentage rises, one probably ends up either trusting all mail from a particular source, regardless of the author, or with a situation quite analogous to “whitelists,” although one that is much harder to trick than the original. Either is subject to attacks and scaling problems.

There is also the risk of abuse by providers who conclude that mail that cannot be authenticated well enough that their users can prioritize it should simply be rejected and who then define the conditions for adequate authentication in terms of a small circle of cooperating mail providers. Even if the types of authentication outlined in Dave’s article are used only as intended, the costs to recipients will rise, perhaps rapidly, over time as percentages of messages bearing authentication information rises and sender authentication and authorization become just one more tool to distinguish probably-desired messages from probably-undesired ones.

Maybe there is not Enough Spam Yet

One of the depressing consequences of the reasoning discussed previously is that perhaps we have yet to see sufficient spam for governments and regulatory bodies to take the spam problem seriously—seriously enough to deploy effective laws and enforcement mechanisms. If spamfighting methods shift the burdens of receiving spam away from those who have the resources to protect themselves they may simply place the spam impacts on others who have fewer resources. That pattern may, in turn, also reduce pressure on governments to take effective action and to do so in a way that would make the design constraints for effective technological approaches clear. If a collection of anti-spam methods have the effect of simultaneously increasing the amount of total spam on the network and of decreasing pressures on societies and governments to take effective action, are they really ones we want to deploy?

Conclusions

This article presents a rather grim view of the future if we continue on our present course. If we fail to examine the actual actions that societies and their governments are willing to take to deal with spam and spammers and to treat those actions and their limitations as design constraints on the technical and engineering approaches, we are likely to continue to see an ever-increasing amount of spam on the network. Spammers will not only adopt technical countermeasures to new techniques but they will also take advantage of their ability to simply increase message volumes (at almost no cost) to counter the effects of those techniques on the percentage of spam that is delivered. It may be time to finally deal with the spam problem as the difficult social issue that it is, rather than permitting societies and governments to continue to believe that a technological “silver bullet” is right around the corner and that no real social or political action, or commitment of law enforcement resources, is needed.