A Brief History of the War on Spam

Nathaniel Borenstein, Chief Scientist, Mimecast

Spam is not a simple problem, and it is not likely to go away, ever. Spam says profound things about the limits of productivity and wisdom. It is the electronic static that limits collective thought.

Although spam gained notoriety as a concept within an email context, as a phenomenon it arises inevitably in any sufficiently rich and open system of communication. From the perspective of the individual, many things are 'spam.' Billboards are visual spam, junk phone call and faxes are telephony spam, unwanted audio at the gas pump is vocal spam, and so on. If a system of communication is open to everyone, it will have spam, especially if it eschews central authentication authorities.

Unfortunately, spammers have only begun to explore the range of options and techniques open to them. Spamming will be big business for a long time. Accordingly, the struggle to separate the "good" communications from the "bad" is becoming a central focus of human interaction with computers. Although spam cannot be eliminated, an intelligent and deep program of spam control has, so far at least, been able to reduce it to the level of a minor nuisance.

Spam is not the kind of problem that can be solved simply, because unlike most software challenges, anti-spam measures are battling active, intelligent opponents. Rather than a puzzle to be solved, the spam problem is best approached as a long term competition. This dilemma can be modeled in terms of game theory as a contest between active opponents. When looked at in that light, it is clear why the spam fighters have to keep building ever more complex countermeasures, with no end in sight.

One of the causes of this seemingly endless treadmill is, perhaps surprisingly, Moore's Law. More of an observation than a law, Moore's Law states that the number of transistors we can put on an integrated circuit roughly doubles every two years. Moore's Law is the magic behind the phenomenal growth in computing power over the last several decades, and without it we wouldn't have smartphones, laptops or lots of other invaluable things. But in the field of spam, Moore's Law is on the side of the bad guys.

Think of it this way: a team of researchers works for two years and cuts the "false negative" rate — the rate at which spam gets through to users — in half. Meanwhile, the spammers need do no research at all, but at the end of the same two year period, they have the ability to send twice as much spam for the same cost, and the net amount reaching the users is unchanged. It's not really quite that simple, because there are factors in play other than sheer quantity, but it shows how the bad guys start out with a significant structural advantage.

The fight between spammers and spam fighters is one of long term dynamic equilibria. The spammers probe for ways to get more spam through, and the spam fighters develop countermeasures. The system is generally in rough equilibrium, but the point of equilibrium can be nudged in either direction. In the fight against spam, we have two general categories of techniques for moving the equilibrium toward less spam: technical and non-technical. The technical measures are primarily the job of anti-spam technology companies and mail system operators, while some of the non-technical measures are primarily the responsibility of any organization that depends on email. We will discuss each in turn.

Technical Countermeasures

Filtering

Filtering messages based on content is the first and most widely used approach to spam fighting. Email filtering can take place at various stages in email delivery. It is used for outbound messages, to make sure your company is not becoming a vector for antisocial messages, and for incoming messages, to protect your users from malicious junk. The filtering can take place on your own mail servers, on servers belonging to a third party such as a cloud provider, on intermediate relays, or even on an email client.

Because spammers are constantly varying their messages, successful filtering depends on regular and timely updates to filtering rules. The distribution of such rules can be complicated, because the spammers could cause a great deal of mischief if they could modify the rules. Most systems that filter email get their rules from a relatively small set of well-known providers.

Although most email filtering uses such centrally managed and automatically distributed rule sets, filtering with fixed rules that are regularly updated, other systems use automated learning algorithms to evolve the rules. Such systems typically require some kind of user feedback, such as a “Report Spam” button, to drive the learning mechanism.

Although content-based filtering was initially quite effective, it has become steadily less so. In fact, game theory and experience both suggest that the attempt to defeat spam by filtering is doomed in the long term. Because the spammers can watch and respond to trends in filtering, they can continually innovate their approaches to get around it, as well as simply send ever more messages.

Authenticating Identity

If spammers would only identify themselves clearly, it would be easy to block spam. This observation has led to a plethora of whitelists and blacklists, all designed to assess the identity of an email sender against his or her reputation, based on prior actions. Unfortunately, it’s not nearly that simple.

Given the widespread desire to blacklist known spammers, a critical part of being a successful spammer is making messages appear to come from someone who isn’t a spammer. In order to avoid a simple name-based blacklist, for example, spammers vary their “From” identities as much as they can, sometimes using dictionary-based names, with results like “Ulysses S. Kleenex.”

As humans, we have a built in tendency to trust things based on their names. Winston Churchill said, "the From field is useless for authentication, but excellent for human engineering." OK, you probably didn’t believe that, but only because Churchill obviously lived and died before ‘From fields.’ Still, the attribution of a message to a name you respect is almost guaranteed to make you at least begin to read it.

For better and for worse, the fundamental design of the Internet makes guarantees of identity very hard to provide. If the entire Internet was controlled by a single company, it would be relatively easy for that company to provide a strong identity authentication of message senders. But it would also be a far less free environment, with the potential of censorship at that company’s whim. And strong authentication could actually endanger the lives of, for example, activists fighting dictators around the world.

Instead, the Internet connects millions of machines, each of which is controlled relatively independently. Each site can choose whether and how to authenticate messages that originate in machines it controls. But such authentication becomes more or less impossible to confirm at other locations in the Internet where those messages end up.

There are ways to fix this problem, but they face strong opposition, which makes them virtually useless. For example, the Internet community has been working for over 20 years to develop person-to-person authentication in email, resulting in email encryption systems known as S/MIME and PGP. These systems have seen stunningly low adoption rates, in part because of their perceived complexity, and in part because people don't seem to want strong authentication most of the time.

In recent years, domain-based authentication has emerged, using standards like DKIM and DMARC, whereby cooperating sites can authenticate messages based on where they originate. This allows sites to make informed judgements about the mail that comes from another given site, and how likely it is to be spam. Messages from "good" sites can be passed through, while others can be filtered more heavily. This system has tremendous potential, but it also substantially complicates the work of anyone operating an email service, and provides an additional motive for outsourcing the bulk of email management.

And, ultimately, as long as a sizable portion of the Internet isn't cooperating in the same authentication scheme, the Internet's infrastructure will continue to permit spam.

Payment Models for Email

Another approach to spam control, widely discussed but rarely implemented, is to try to change the economics of spam by imposing a cost that discourages rampant email abuse. With traditional postal mail, the quantity of junk mail is limited by the cost of postage. Senders have an incentive to try to only send mail to people who might want it, rather than waste money on those who clearly don't. Such incentives clearly don't exist for email. Imposing a payment model would be a way to change that.

While this approach has a certain appeal to many, it also has many problems. First, there is widespread resistance to the idea of paying for "good" email. The fact that email is essentially free to send is widely seen as a major factor in its favor, so that many would be unwilling to give it up – even to eliminate spam.

However, there are alternatives to the "postal model" of charging for spam. The most interesting of these link money and authentication practices, and only charge the sender when that authentication fails.

For example, Marshall van Alstyne has proposed a system called "attention bonds," in which a message sender would entrust a relatively small amount of money to a trusted third party. This money would serve as a "bond" to be paid out if, and only if, the recipient of the message flagged it as junk. If a recipient claims the money, the sender has a powerful incentive not to send him any more mail, while if a recipient finds a message useful, he has an incentive not to claim the money, or else he would be cut off from future mailings. "Good" mail senders would have most of their mail delivered free, but anyone who tried to deliver junk mail in bulk with this system would incur large costs that would make spamming unprofitable.

In another example known as 'charity stamps,' attempted previously by companies like Yahoo!, senders link their email to charitable donations. For each email sent, the sender needs to demonstrate that a certain amount of money has been given to charity. For most people and corporations, this would be under the amount they give to charity anyway, so the mail would be incrementally free, while spammers would either have to stop spamming or begin to give the world some social benefit each time they spam.

To date, none of these systems has found widespread adoption, but there is a chance the concept will find

renewed interest in an age when email-based attacks and annoyances continue to escalate.

Non-technical Countermeasures

Enforcement

Everyone hates spam — it's one of the most agreed upon propositions in the world. Inevitably, there are brave politicians who work to make it illegal. Of course, if making something illegal automatically got rid of it, death would be illegal by now, too. Nonetheless, a tangle of laws has grown up around spam that is very complex and even occasionally useful.

The first thing to note is that spam, like the Internet itself, is a global phenomenon. You are as vulnerable to a spammer in Russia as to one next door. In the absence of a global police force, enforcement depends on a high degree of cooperation — first in tracking down the spammer, which can be a very complex technical process, and then in bringing him or her to justice. Because spammers try to cover their tracks, this might involve getting subpoenas to access ISP records in a half dozen countries or more – and smart spammers make an effort to work through countries that cooperate badly with each other.

It generally takes months to find a spammer, and then the legal process of indictment and extradition begins. The amount of effort is huge, so police are generally reluctant to take on such cases. A few agencies, including the FBI in the U.S., have a broader reach and do pursue spammers occasionally, but given the cost required, they pursue only the worst offenders.

That's where enforcement is sometimes useful: finding and shutting down the biggest spam offenders. When you hear that a spammer has been arrested, there are generally enormous numbers of spam messages attributed to them. If they hadn't sent spam by the billions, they wouldn't be worth the effort to catch. Culling the very biggest spammers is yet another mechanism in the fight against spam, and pretty much the only reason anti-spam laws aren't completely ineffective.

Even more dubious are the "do not email" registries that are scattered about the net. Even when they're not owned by the spammers themselves, they often function as a convenient way for spammers to get more email addresses, therefore having exactly the opposite of the intended effect.

Education

Ultimately, the best hope for beating spam and other malware is to train users not to be fooled. Most malware depends on fooling users into clicking on the wrong link, opening the wrong attachment, or following the wrong instructions. The better educated users are, the less often there will be negative consequences from malware. Both organizations and email providers can provide a program of regular messages that give clear information and examples of what not to do. Instructions to users should be short, clear, and infrequent; they should receive no more than a few email updates every month.

Ideally, safe email habits will eventually seem like common sense to the average user – like locking their front door before leaving the house. Since email is not going anywhere anytime soon, despite the overeager predictions of some, educated email users are essential to the future of a functioning Internet.