What it is, how can it affect us, and how to deal with spam.

Adalberto Zamudio
What it is, how can it affect us, and how to deal with spam.

Abstract:

This paper would explore the problem of spam. It would explore its nature and also the reasons why it is a security risk. It would also describe some of the techniques that are currently used to deal with spam including basic technical and policy methodology. This paper intends to be a general-information source to the problem of spam and also provides information of where to get more detailed information about the products that and techniques that describes.

Part 1: Background: What's Spam?

The main purpose of this paper is not only to analyze the reasons why Spam could be considered as a security risk, but also the different ways to deal with it. Before starting the discussion about it, it is important that the reader understands what spam is. The main reason to set this background is to aid in the differentiation between what is spam and what is not, what are the requirements that an e-mail must have in order to be considered spam; This, in order to have a clear idea of what constitutes spam and then being able to use that concept to understand how could an e-mail be considered a security risk and, moreover, what can be done about it.

There are different ways to establish the concept of Spam, and this paper is not trying to ignore the fact that using a definition is not the only way to do this, but, due to the fact that found the concept is only to bring into being some background and then start the discussion of the main topic, a definition, and it's appropriate analysis, should be an adequate amount of information for the scope of this document.

There are many different definitions of Spam, when searched for “spam definition” (inside quotation marks) the search engine Google (www.google.com) returned 2,390 results. Different definitions look at different aspects of spam, and some are based on specific characteristics. This paper will use a definition by Mail Abuse Prevention System, LCC (MAPS) for reasons justified later.

MAPS distinguishes an e-mail message as Spam, if the following 3 conditions hold:
(1) The recipient’s personal identity and context are irrelevant because the message is equally applicable to many other potential recipients; AND

(2) The recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent; AND

(3) The transmission and reception of the message appears to the recipient to give a disproportionate benefit to the sender.¹

This paper is using the definition by Mail Abuse Prevention System mainly because this classification of Span doesn’t make specific reference to commercial e-mails (as most definitions encountered while researching this topic do) as a source of Spam. MAPS require that three conditions on an electronic message hold in order to consider it spam. These three conditions are, as mentioned before, both general and specific enough to encapsulate all the different instances where Spam is present. Due to this fact (that is, the fact that the definition is both specific and general) this definition requires further examination and the use of examples.

The first condition of the Mail Abuse Prevention System’s definition requires that the recipient of the e-mail message be irrelevant, that is, that the SAME e-mail could be sent to a different addressee with minimum changes; In this case the minimum changes could be something like the name on the header of the e-mail: change Mr. Smith for Mr. X and the condition holds. Every user of e-mail has roughly the concept of spam as a number of messages that appear on their mailbox with different topic that are sometimes of interest and sometimes are not. The definition of MAPS generalizes this when it says that this message could be as well intended to be for many other recipients; that is why, sometimes spam mail could be relevant for the reader and sometimes not. This definition generalizes the concept since it is making reference to a template e-mail sent to the user’s mailbox.

The second condition for an e-mail to be considered Spam, is probably (based on the other definitions of Spam encountered while researching this topic) the one that would make reference to the most common idea of Spam, that is that the recipient didn’t asked (or gave permission to the sender) to send the specific e-mail. The definition goes beyond this when it requires not only the permission to be verifiable (that could be on an electronic or hard form) but also that permission, in the case of it being present, to be revocable. As mentioned before this clause will be similar to most of the ordinary definitions of spam and to the general knowledge ones since it will make reference to the mailboxes flooded with unsolicited e-mails.

The second condition also provisions the fact that when a recipient enquires about the reason why the specific e-mail was sent to his or her mailbox, the sender could provide a request or authorization form; moreover this authorization form should be revocable whenever the recipient decides either that the e-mail is

¹ Mail Abuse Prevention System, L.C.C.
not what he wanted, or that its content is not relevant any more. This condition is exists in order to make room for legitimate identical e-mails that get the “red flag” on the first condition, but then, assuming that the recipient asked for this e-mail (say, a newsletter) to get a “green flag” and then not to be considered as Spam AS LONG as the e-mails could be stopped.

Mail Abuse Prevention System’s third condition is key to separate legitimate e-mail from Spam. This condition is necessary since there could be some instances where template e-mails could be sent to a mailbox where no option to cancel is available, but that message could not be Spam. According to the definition, there should be a “disproportionate benefit to the sender.” An instance where you signed up for an e-mail about the weather in your city (that is, this e-mail is not specifically directed to you since the exact same e-mail could be sent to any other person living in the same city), and there is no option to cancel (this is, the authorization to send the e-mails is not revocable) might not be Spam, since the content of the e-mail just contains information about the weather in your city. In this case, the recipient is the one that is getting most of the benefits out of receiving the e-mail.

Looking closely to the rule, if this e-mail about the weather contains a disproportionate number of advertisements (compared to the weather information present on the e-mail) then, since the other two conditions for an e-mail to be Spam hold, this e-mail could be considered as Spam, that is, this e-mail does give a disproportionate benefit to the sender. As mentioned before, the third condition is the key to separate legitimate e-mails that, according to other definitions could be classified as Spam, from actual Spam.

It is also important to note that Mail Abuse Prevention System’s definition also requires that all conditions hold. This is, the fact that someone receives an unsolicited e-mail is not enough to consider it as spam. Looking closely at the conditions, they cover all the different instances that most of the other definitions consider as spam, for example the one by The Government of the Hong Kong Special Administrative Region’s Info Sec: “Unsolicited email, often of a commercial nature, sent indiscriminately to multiple mailing lists, individuals, or newsgroups.” The difference is that by considering all conditions, they take into account the intention of the sender and they are not concerned about the content of the e-mail per se, but the intentions and the overall meaning of it. (not to mention the fact that even solicited e-mail could be considered as Spam as long as there is no option to cancel and the other conditions hold).

Now that the background is set up with the definition and analysis of what constitutes (or not) Spam, the next section of this paper will deal with analyzing the reasons why Spam could be considered a security risk and also analyzing
what can be do about it, that is, the different was to not only manage the risk, but also to counterattack Spam!

Part 2: Exploration: How much does it costs?

This section will be concern with analyzing how much does spam costs, and therefore, why could spam be considered as a security risk (since the fact that spam exists is costing resources). Spam’s costs could be seen in different levels, and this paper will deal with what it considers the two main types: direct and indirect costs. The first kind of cost would include all the costs incurred by a person or organization by receiving spam. The second one involves the social engineering and scams that happen through spam and the costs associated to those activities.

Shoshannah Forbes analyses the “direct, unhidden costs for the privilege to get junk” and identifies different expenses. The most obvious one is the time spent on the Internet while downloading the Spam. In this particular case Forbes stands out the fact that her connection to the Internet is via dial-up to an ISP, and then calculating the cost of downloading spam in quite easy to calculate (the time spent on the activity times the unit cost of the connection). She also identifies the cost of digitally storing spam. It is important ton note that even though hard drives are becoming cheaper, and therefore the costs of digitally storing one megabyte of information becomes cheaper, it is becoming a common practice to include banners with images on e-mails or multimedia which are larger files and then, the costs of digitally storing e-mails could be considered as remaining constant.

Even though at first sight these direct costs might not appear to be substantial, it is important to consider that there are some 600,000 people connected to the Internet, and Global Reach estimates a 940,000 by 2004. A study by the Stanford Institute for the Quantitative Study of Society points out that “E-mail is by far the most common Internet activity, with 90% of all Internet users claiming to be e-mailers.” That’s about $270,000 spent on direct costs of spam worldwide (and about half a billion dollars by 2004!) And that’s using what this paper considers as a conservative calculation, that is Forbes “$0.53 per month” for the average user based on 20 spam e-mails per day.

On top of the direct cost, there are some hidden direct costs related to spam. When calculating the direct cost of spam we are counting as a cost the time spend on an internet connection when downloading spam, but as Henry Butz

4 Forbes and Butz
5 Global Reach
6 Global Reach
7 Borom
8 Forbes and Butz
points out, “how do you know they were spam unless you read them?” Following Butz’s line of thinking, it is important to consider, then, other hidden factors that might be considered as direct costs as well. That statement alone should make us think of the cost of the time, and that is not only the cost of being connected to the Internet for that lapse, but also the cost of people’s time, of spending that time downloading and reading the spam when they could be doing something else. David Ho, an Associated Press Writer reported that Enrique Salem, president of Brightmail, said to the Senate Committee on Commerce, Science and Transportation: “Spam costs U.S. businesses $10 billion each year in lost productivity.”

The reader should now have an idea about the direct costs of spam. This might be of special interest to businesses, where productivity could be affected by time and resources wasted on downloading and reading spam. Especially if we consider the amount of spam that is received at businesses. According to ZDNet “30 to 50 percent of workplace e-mail is spam.” It is the costs associated to spam what makes it (at this level, to start) a security risk. It is the potential that spam has of making someone to lose something (time, money) that makes it dangerous.

So far, this paper has only considered the direct costs resulting from receiving spam, but it is also important to consider the indirect costs that could result out of replaying, or following the instructions on spam e-mail. The second part of this section will deal with those indirect costs, since it considers them critical to understand the impact that spam could have.

“The same scams that have been conducted by mail and phone can now be found on the World Wide Web and in email, and new cyberscams are emerging.” Scams are not something that came with the Internet, but it is a fact that the same way that using the internet can aid scammers to reach more people than more traditional (post, phone) ways at a fraction of the cost (since e-mail is cheaper than sending a letter in the post). According to the National Fraud Information Center, Internet frauds overall losses have gone form US$3,262,834 in 199911 to US$14,647,933 in 200212, that’s over 448% in just 4 years!

It has to be said that general Internet fraud is out of the scope of this paper, but even though “Web sites are still the most common way that consumers are solicited for fraudulent Internet offers (…) the statistics reveal an increase in the number of initial contacts made by con artists in emails,”13 What this is telling us is that although the actual scam doesn’t necessarily happen ‘via e-mail,’ e-mail,

---

9 Forbes and Butz
10 National Fraud Information Center, “About Internet Fraud Watch.”
11 National Fraud Information Center, “2001 Information Fraud Statistics.”
12 National Fraud Information Center, “2002 Information Fraud Statistics.”
13 National Fraud Information Center, “2001 Information Fraud Statistics.”
and specifically spam, is the vehicle used for the first approach. If we refer to the National Fraud Information Center statistics from 2002, we can see that only 6% of the reported frauds happened over e-mail (that is, people sending their credit card numbers or other form of payment over an e-mail). This fact is understandable, since must people know that e-mail is not a secure medium; the problem is that when the when a website uses encryption to transmit data doesn’t makes it ‘secure’ against the corrupt intentions of whoever is at the other side of the computer. It is by sending spam with links to a specific web site (or telephone numbers) that those hoaxes start and that’s why they should be included in the calculation of the risk associated with spam.

So the danger about spam is not only that it wastes resources, but also that it could lead to higher losses, not only of money, but also of time, and if we take into account the discussion above, time could represent an important resource (and affect productivity). Just to illustrate the point, we can use the ‘Nigerian money’ hoax example. “It comes, often quite regularly, from an alleged former dignitary of the Nigerian government. The typical storyline is that the sender has stashed away a huge wad of cash, but needs a foreign bank account through which to funnel it.” What happens next varies bit, but mainly consists of the Nigerian official offering to share the money in exchange of help to get the money out. People sending their bank details, and as a consequence they have their bank accounts cleared or used in other scams. There have been instances where the alleged Nigerian officer has the money on a safe box and needs to pay the bank to access it, sometimes to pay a over-due loan, after people give the money to open the box they never see the Nigerian official, or their money, again!

As the reader can note, spam could be more than irritating unsolicited e-mail. As reviewed above, spam has direct and indirect associated costs that can go beyond the irritation that one can get out of getting 20+ irrelevant e-mails in their mailbox, spam is costing businesses money, and while doing that it is affecting the economy. Spam is also an instrument that can be used as a starting point for other hoaxes that can lead to greater losses and that is why spam is a security risk, due to its catastrophe potential. Now that we have a good idea of what spam is, and the cost of it, the logical next step is to know how to deal with it. The next section will explore different options available.

Section 3: Discussion: How to deal with Spam?

The following section of this paper is, intended to give general details about the most commonly used anti-spam methods. There will be two sub-sections; the first (the one to which this paper will dedicate more time) will deal with the specific techniques that exist in order to deal with spam. The second will explore the different ways that an end-user can fight spam. The two sections are not exclusive, even though they tackle the problem from two different angles, the two can complement each other, since there is no perfect method to fight spam.
Since spam is so frequent these days, it is not a surprise to find a significant number of companies and methodologies to fight spam. When searched for “ant-spam” (with quotation marks) on the search engine Google (www.google.com) returned over 1,100,000 results. Before continuing and going into more specific information, it has to be noted that, as the National White Collar Crime Center points out “There are several prevention methods for individuals to implement in order to avoid Internet fraud victimization. The most important step is to stay informed about current crime trends.” There could be millions of prevention systems, but if the users are not aware of them and they don’t use them, there is no way that the spam is going to stop. The only way to win the battle against spam is to be informed and, based on that information, take the necessary actions to defeat the spammers.

Since being informed is the best way to combat spam, there are a series of best practices that should be followed by the industry in order to prevent spam. These practices are not going to stop spam by themselves, but the idea is to make it harder for the spammers. There is a group called Bestprac.org that specializes on creating the mentioned best practices. They have divided their practices on 20 different groups:

- Backbone Providers & Bandwidth Wholesalers (BPR)
- ISPs (ISP)
- Web Hosting Services (WHS)
- Mailing List Hosting Services (LHS)
- Search Engines & Directories (SED)
- Free Web Email Services (FWE)
- Free Website Hosts (FHS)
- Phone/Fax to Email Services (P2E)
- Web/Email to Fax Services (W2F)
- Third Party Script Hosting Services (3SH)
- Other Free Web Services (FWS)
- Affiliate / Associate Program Managers (APM)
- Affiliate / Associate Program Affiliates (APA)
- E-zine & Mailing List Publishers (MLP)
- Browser Software Developers (BSD)
- Email Client Software Developers (ECS)
- Email Server Software Developers (ESS)
- Media Buyers (MBY)
- Domain Name Registrars (DNR)
- Webmasters / Web Designers (WMD)
- And other industry Participants.

---

14 National White Collar Crime Center
15 Bestprac.Org
The idea of the best practice principles is mainly to inform the different groups and give them the appropriate ‘rules’ in order to “to promote and encourage responsible and ethical practices throughout the Internet industry, for the prevention and eradication of email spam.” As the reader can see form the name of the rules, and now that the basic concept of what those best practices are, these different groups are mainly the different stakeholders that make the internet function, and that’s why it is important to ask them to follow the best practices. These principles are, by no means, the complete solution to the problem, but if all the different groups followed them, it would be very difficult for spammers to continue, at least on a profitable way, and then they would be forced to stop.

To go into more detail about each of the principles for each of the categories is beyond the scope of this paper, nevertheless is important to give a general description of the kind of principles that we’re talking about here. The principles are mainly recommendations of different ways to protect e-mail addresses that could be used as a ‘mail list’ by spammers, also, ways to ensure that spam could be reported easily and that the appropriate measures can be taken on a fast and efficient way. There is special emphasis on protecting e-mail addresses since “email address harvesters, crawlers or spiders, are quite probably the single most insidious manner by which spammers collect the email addresses of innocent victims.” As mentioned before, if the principles are followed it would be harder for spammers to get hold of ‘mail lists’ to send their e-mails to.

It has to be said that Bestpract.Org is not the only company with best practices, it was selected to appear on this paper since, to the consideration of the writer, it provides the most complete list of pest practices for different all the groups at the time or writing, and all the information can be found on a single webpage (www.bestprac.org). The Internet Society also produced some suggestions on “How to Advertise Responsibly Using E-Mail and Newsgroups” and other topics, but as said before these are just suggestions, and does not constitute an Internet Standard. Mentioned that, The London Internet Exchange has a Subcommittee that deals with Unsolicited Bulk Messaging and they have produced a list of recommendations as well (for more information see the references at the end of this paper). The list of companies or organizations that produce guidelines could continue, but now the reader should have a good idea of what constitutes the best practice guidelines, and that’s as far as this paper is going to go.

Besides best practice guidelines, there are other methods of fighting spam. One of the most effective ones is the use of a blacklist. “A black list is a list of email addresses that you never want to receive email from.” So a basic filtering combined with the database of the blacklist could stop some spam, i.e. a rule

---

16 Bestprac.Org  
17 Gavin, Eastlake and Hambridge  
18 Privacy Labs
saying that if an e-mail sent from any of the e-mail addresses on the list is received it should be deleted (or sent to the “junk mail” folder as appropriate). The next logical question is how to get the e-mail addresses of the ‘bad guys’ (the spammers).

There are a number of databases of known spammers. It is a good idea to have this kind of databases since “90% of all spam received by Internet users in North America and Europe is sent by a hard-core group of under 200 spam outfits.” As mentioned before, there is not a unique database with all the bad guys, and as expected there can be mistakes. In most of the blacklists there is a section for people that have been blacklisted and think that this is a mistake. These lists are updated on a regular basis, regularly more than once a day around the clock from sites worldwide.

In order to give some practical information to the reader about where to find the blacklists mentioned above, and this is a good moment to say that in order top access some of this lists a fee has to be paid, this paper will include a few blacklists and the Internet addresses. The Spamhaus Black List (SBL) at http://www.spamhaus.org must be on top of the list since it is, at the time of writing, one of the most comprehensive lists that sends updates on an hourly basis (with mirrors around the globe). There’s also Uceb.Org at http://www.uceb.org and many others. Some companies and Universities have their own blacklist, and they all function in a similar way, the difference might the people on the blacklist.

Another very popular method are the user-personalized blacklists. This methodology allows the user to select which specific senders (in the form of an e-mail address or domain) are not allowed to send e-mail to them. The rest is pretty much as a traditional blacklist, if an e-mail from someone in the list is received, this e-mail is discarded, or sent to the “junk mail” folder, as selected by the user. This is, by no means, a perfect technique, but it allows the user to take control of what kind of e-mails to receive. At the same time, it requires more of the user’s time since he or she has to set individual filters each time spam from a new source is received.

There’s also the concept of a whitelist; is pretty much the opposite of the blacklist in the sense that only the e-mails from the people on the list would be received and everything else will be deleted. “Setting basic filters to accept email only from a finite list of approved senders, deleting virtually everything else, would eliminate most spam. It would also, however, delete legitimate email as well.” A simple whitelist is not a solution unless the user decides that he or she doesn’t want to receive e-mails form new people, which is something that even though might be appropriate in some circumstances, will not fit most of the cases.

---

19 The Spamhaus Project
20 Mailshell
There is another way to deal with spam; there are some content filters in the market. The idea is that they scan the received e-mails and, depending on the words that are included, the e-mail receives a ‘score.’ With a simple filer, if an e-mail message has more than a certain number of points (this could be set by the user to tighten or widen the filter). An example of this kind of filers will be SpamCop at [http://www.spamcop.com](http://www.spamcop.com). The problem with this kid of filters is that they produce a number of false positives, “resulting in tagged mails which are not actually SPAM at all.” This could be a problem as well since the result is that users have to scan their ‘junk mail’ folder in search of non-spam e-mails, and that’s a big waste of time!

There is also Tagged Message Delivery Agent (TMDA) that “combines a "whitelist" (for known/trusted senders), a "blacklist" (for undesired senders), and a cryptographically enhanced confirmation system (for unknown, but legitimate senders).” This application seems to use the best of the three techniques. No to the blacklisted addresses, yes to the whitelisted ones and then if only content-filters the ones that are not on either list. This sees to be a good technique. For more information of other anti-spam tools (and a ranking), the reader can go to PCMagazine’s anti-spam tools review at: [http://www.pcmag.com/category2/0,4148,4795,00.asp](http://www.pcmag.com/category2/0,4148,4795,00.asp)

Alternative strategies to deal with spam have been proposed. For instance, François-René Rideau proposes “a plan to eliminate the bulk of unsolicited commercial messages from electronic mail services by shifting the cost of spam back to bulk emailers.” Rideau’s idea is that in order to send e-mails people could attach a ‘stamp’ or that would cost them some defined amount of money in order to show their “good faith.” Following basic economic theory spam should reduce, that is given that Rideau’s idea actually takes place. He suggests that a way to push the method is to blacklist the senders that don’t comply with the scheme. There are a some technical and policy challenges with this idea, but it shows that some theories are being develop that think outside the box. This reflects the fact that the market is reacting upon a problem that is getting more serious with time.

So far, the techniques described are pretty much user exclusive, that is, they try to simplify the user’s role. This is not surprising since one of the reasons why we want to stop spam is because it wastes the user’s time, and affects productivity. Even though that’s a valid reason, this paper will still explore, briefly, the different things that can be done taking the user as an active player on the fight against spam. It is important to clarify that the whole concept relies on the assumption that best practices are followed (for more information see above). The whole

---

21 McMaster University
22 Tagged Message Delivery Agent.
23 Rideau
24 Rideau
idea goes in the lines of Rideau’s concept of making the spammer’s pay for sending the e-mails, but in an indirect way.

The user can make the spammers pay by reporting spam to the ISP, which, hopefully, will follow the best practices and have an easy way to report spam (abuse@the-isp.com) and will have guidelines to deal with it. The spammers will have to invest time and money to get a new ISP. If the e-mails is of a commercial nature, a simple e-mails to the company stating that you received a spam and that you’re never going to do business with a company that uses this kind of advertisement should trigger the appropriate alarms (well, after a few hundred of users do that) such that the companies stop using spam because that would affect their image, and profits! This might end up forcing the user to spend a lot of time which creates a problem trying to solve a problem, but the idea of the payoff in the long run, that is, stopping spam to get into their mailbox, should be a good incentive for them to send that time.

Part 3: Conclusion

The reader has seen what spam is, how can it affect us and how much could it cost, and how could spam be a security risk. It has also seen different techniques to deal with spam. It is believed by the writer of this paper that there is not a single method that would get rid of the problem. It is the combination of methods that would, at the end of the day, be able to eliminate spam. It would be the combination of technology and policies that could have a more effective impact on the problem. There is also the legal since, but, at least to the time of writing, in the United States “Spam legislation has not yet been enacted at the federal level. Several spam-related bills were introduced in previous sessions of Congress, but none were enacted.”25 And there aren’t any specific laws in other countries that would address the spam problem in a direct and effective way. Nevertheless, the legislations has started, and combined with the other methods (some of them described above) an end to spam could be reached, and therefore, the end of the security risk that spam represents.

25 Sorkin, David.
References


### Upcoming SANS Training

Click here to view a list of all SANS Courses

<table>
<thead>
<tr>
<th>Program</th>
<th>Format</th>
<th>Availability</th>
<th>Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>SANS OnDemand</td>
<td>OnlineUS</td>
<td>Anytime</td>
<td>Self Paced</td>
</tr>
<tr>
<td>SANS SelfStudy</td>
<td>Books &amp; MP3s OnlyUS</td>
<td>Anytime</td>
<td>Self Paced</td>
</tr>
</tbody>
</table>