

# On Unsolicited Junk E-mail, a/k/a “spam”

by J. Harmon Grahn

## A Short Horror Story

Until recently, I have considered so-called “spam,” or unsolicited junk e-mail, to be what you might call a “minor irritant,” which I customarily chalk up to one of the “costs” of the convenience of e-mail; and just shrug, and massively delete (taking care always not to lose the occasional genuine message I actually want to read).

For the second time recently, however, I have had this matter of “spam” pushed directly “in my face” – when I downloaded my e-mail, and was inundated by a flood of hundreds if not thousands of “Postal Notices” to the effect that messages I had sent to e-mail addresses all over the world were for various reasons undeliverable. Some of these automated replies were written in Chinese and Cyrillic, for Pete’s sake!

I had never sent any such messages, of course. Some “industrial-strength spammer” had evidently launched a massive distribution, *using my e-mail address as the Sender*, and so I was receiving hundreds of notifications of failure to deliver to bogus addresses in the unknown spammer’s distribution list. As a result, I suddenly became *motivated* to look into this matter of “spam” a bit more closely than I had before, and I think it may be fruitful to share with you some of what I have learned. This may be considered follow-on, in significantly broader scope, and greater depth, to a “Dear Friends” letter I posted here 04/11/2007 21:39.

It is doubtful that anyone who uses e-mail has not been forced to contend, one way or another, with so-called “spam” from persons unknown, concerning matters of little or no interest to the recipient. In fact, *junk* is by no means too harsh a term for this perpetual flood of often noxious, tasteless, offensive content that arrives daily and hourly, 24/7, in our often indiscriminately receptive e-mail in boxes. It is like a perpetual snowfall of tiny, light flakes which seem weightless and inconsequential when viewed singly – for they are after all only a few bits and bytes of weightless digital data which may be instantaneously deleted at the command of a button. Yet in their aggregate they accumulate as massive glaciers that could imaginably overburden the global Internet, provoking the digital equivalent of a latter-day Ice Age.

The massive tide of “Postal Notices” I received are not themselves junk e-mail; they are legitimate responses to unsuccessful attempts to deliver mail, and when someone uses your e-mail address as the return address on a massive spam distribution, guess who receives the notices about the undeliverables? Bingo! You do.

## Seeking a Solution

Has anything like this ever happened to you? If not, I can tell you, It’s more than a “minor irritant;” and so when it happened to me for the second time in as many months, I started casting about to see if there’s anything I can do about it. So I wrote one of the local Net Gurus here, describing what I have just related, and he replied, in part, as follows:

Well, this sounds to me like bounces from spam. Spammers always forge the return address on email they send. Often they will use addresses on their list as reply-to addresses, so as to cause bounces to go somewhere other than back to them, in that case to others they intend to spam. In this way the spam may end up bouncing somewhere they want anyway, so that it bypasses spam filters and someone reads it.

Anyway, the bad news is that there is nothing that can be done about other people using your email address as a reply-to address. You can do that, too. In fact, it's possible for anyone to send email with the return address of `president@whitehouse.gov` for example, and if the message doesn't reach its recipient, it will bounce back to that address, even if the message was not sent from that address. Your spam filters are set as high as they will go, and it's very difficult to block such messages ["Postal Notices"], as they are coming from a mail system and aren't technically spam. Also, `harmonhouse.net` is not hosted by us, so email sent to that system is not in our control.

I notice that your email address `somebody@somewhere.com` [substituted here for my actual e-mail address] is in a plain `mailto:` link on your site, and that can be harvested by spammers easily and used. The longer your email exists in plaintext on your site, the more lists you will end up on, and the more spam you will get. This is very predictable. There are ways to obfuscate your email address on a website to discourage harvesting, but it's a bit beyond the scope of this discussion.

I replied that I had already taken some measures to obfuscate my e-mail address, specifically by replacing the punctuation marks and "special characters" (:, @, and .) in the posted e-mail address with their XML code equivalents, so it comes out looking like this:

```
mailto&#58;somebody&#64;somewhere&#46;com
```

and suggested possibly substituting XML character codes for more of the characters in the address string. My informant replied,

That might work. What I see most often these days is Javascript array and reassembly of the address, but there are many techniques. You can Google for that.

```
http://www.google.com/search?q=javascript+email+obfuscate
```

I followed his suggestion, and spent some time browsing the Google search results. The consensus seems to be that there is not a uniform consensus regarding spam, and measures one might take to discourage it; but it at least made sense to me that *if you post an e-mail address in plain text on a Web page, you're just asking for spam*, because they send "spambots" out roaming the Net, constantly seeking e-mail addresses to add to their vast databases. These are programs that search for tell-tail strings of characters, such as "mailto:", and copy associated strings as potential e-mail addresses to be added to spam lists.

I speculated, initially, that just replacing the punctuation marks may not be sufficient to disguise an e-mail address so the 'bots won't harvest it; and so I set about fixing up a version of my e-mail address in which every character is replaced by its corresponding XML code; and then removing my explicit e-mail address from the link, replacing it with something uninformative, such as "contact" as the active link.

The resulting code string looks, in part, something like this:

```
&#109;&#97;&#105;&#108;&#116;&#111;&#58;&#104;&#97;. . .&#110;&#101;&#116;
```

...which doesn't look very much like an e-mail address in the source code, and doesn't disclose my e-mail address on the HTML page; yet it works just fine, and if you click on the link [in the HTML version of this file], you can send me an e-mail. That would be nice; I'd love to hear from you – but *please be inventive* in composing the content of your **Subject:** field. E-mails from addresses not already known to me, bearing a generic **Subject:** “Hi!” or “Your Website” will be summarily flagged as “Junk” and will not be read. You'll have to convince me somehow that your message is of personal interest to me, and not the work of a parasite.<sup>1</sup>

Anyway, for me, that little change involved a lot of files, because my e-mail address is liberally sprinkled among my many pages, and I want to encourage visitors to e-mail me if they wish. I didn't know at the time if my measures were sufficient to diminish the volume of spam I receive daily; but as I wrote to my informant, “I'm at least sure the coding changes I have made will not produce a worse result than before.”

## Results

Not knowing exactly what to expect, I put these measures into effect on my Website 25 April 2008 – and watched with keen interest the spam traffic addressed to my in box. Additionally, I implemented the e-mail filtering measures available through my e-mail client, Mozilla Thunderbird, which I will elaborate further in a bit. After two full weeks, it is evident that “the coding changes I have made did not produce a worse result than before.” Rather, there seems to have been a noticeable drop in spam traffic to my e-mail address since I removed its cleartext representation from all of my Website pages; and the flood of “Postal Notices” informing me of undeliverable spam I never sent have dwindled to virtually nil. That doesn't mean it can't happen again, because such inundations are episodic, and only occur when some spammer happens to pick my e-mail address out of his (probably huge) list, and uses it as his return address for a particular spam distribution. Nevertheless, I am gratified that the immediate flood of “Postal Notices” has virtually ceased, at least for the time being.

I had hardly dared hope for such immediate results, because my e-mail address had been exposed virtually naked to the world on my pages for the past several years; so I figured that by now I must be on so many spam lists that I'll never get clear of them. It is evidently true, however, that as found in a six-month study I'll discuss further in a moment, “E-mail addresses harvested from the public Web appear to have a relatively short ‘shelf life;’”<sup>2</sup> with the happy result that ending Website exposure of one's cleartext e-mail address is typically followed by a significant drop in the volume of spam addressed to it – as has been my experience over the course of the past two weeks.

## Corroboration

The 2003 six-month study by the Center for Democracy & Technology cited in footnote 2 confirms that even simpler measures than I have taken, as described above, have proven effective in diminishing the volume of spam received by e-mail addresses so disguised on the Net. Even such a transparent stratagem as disguising “somebody@somewhere.com” by representing it instead as “somebody at somewhere dot com” has proven effective in defeating the “spambots.” Moreover, **97.4%, or 8,609 out**

---

1 Not that there's anything necessarily “wrong” with being a parasite, mind you. Perfectly honorable profession. It's just that I do not welcome e-mail from them. If you're interested, see *Symbiosis and Predation in Overview and Synopsis of Metaconsciousness: Mythology for a Post-Civilized World*, V 5.1.0 for elaboration. [[harmonhouse.net/fdl/overview.html](http://harmonhouse.net/fdl/overview.html)].

2 Why Am I Getting All This Spam? Unsolicited Commercial E-mail Research Six Month Report, Center for Democracy & Technology, March 2003 [[www.cdt.org/speech/spam/030319spamreport.shtml](http://www.cdt.org/speech/spam/030319spamreport.shtml)].

of a total of 8,842 unambiguous spam e-mails received during the six-month test were prompted by e-mail addresses posted in cleartext on Web pages. The remaining 2.6% of spam messages received were prompted by:

- e-mail addresses posted to USENET newsgroups = 110 (1.2%);
- opt-out Web services = 82 (0.9%);
- unapproved sharing of e-mail addresses = 25 (0.3%);
- e-mail addresses disclosed on Web-based discussion boards = 15 (0.2%);
- e-mail addresses used to register a domain name = 1 (0.0%).

In summary, the following conclusions were reached in the Center for Democracy & Technology report:

1. E-mail addresses harvested from the public Web are frequently used by spammers.
2. The amount of spam received by an address posted on the public Web is directly related to the amount of traffic that Web site receives.
3. E-mail addresses harvested from the public Web appear to have a relatively short “shelf life.”
4. Addresses posted in the headers of USENET messages can receive significant spam, though less than a posting on the public Web.
5. Obscuring an e-mail address is an effective way to avoid spam from harvesters on the Web or on USENET newsgroups.
6. Sites that publish their policies and make choice available to users generally respected those policies.
7. Domain name registration does not seem to be a major source of spam.
8. Even when an e-mail address has not been posted or shared in any way, it is still possible to receive spam through various “attacks” on a mail server.

The entire report is available on-line [at [www.cdt.org/speech/spam/030319spamreport.shtml](http://www.cdt.org/speech/spam/030319spamreport.shtml)].

## Additional Solutions

I conclude that the method, among others, that I implemented for obfuscating my e-mail address on my Website, described above in **Seeking a Solution**, is a simple and effective way of excluding one’s e-mail address from the largest single resource for spammers seeking e-mail addresses to harvest. Even my previous method of obfuscating only the punctuation marks and “special characters” (`mailto&#58;-somebody&#64;somewhere&#46;com`) would probably have been effective – *had I not posted my e-mail address in cleartext in my links to the obfuscated code*. This oversight, I deem, is the primary means by which my e-mail address became so accessible to so many spammers. It is an oversight I have corrected, and experienced an immediately noticeable decline in the volume of spam I receive daily. This is not a solution that may be expected to insulate one perfectly from receiving spam messages from any source, and I am skeptical there is such a solution; but it seems to work at least in part, and it is not the only means at the disposal of those who wish to diminish the volume of spam they receive.

Earlier, I mentioned the filtering capabilities of my e-mail client, Mozilla Thunderbird. It has a “learning” capability, inasmuch as it takes note of the messages I flag as “Junk,” and is able to take the

same initiative when similar messages, or messages from the same source, are received in future. It also “learns” when I flag a message as “Not Junk” which it has incorrectly “guessed” to be “Junk.” I corrected it a time or two when it made errors of this kind, and it has not repeated them. It doesn’t flag all incoming spam as “Junk,” but I haven’t disagreed with it for quite a long while now about any message it has flagged as “Junk.”

I have instructed it to move immediately all incoming messages it flags as “Junk” to a Junk folder I have set up for the purpose, and to move all such messages accumulated in the Junk folder that are more than four days old to the Trash folder. There I have one last chance to review them before “Emptying the Trash.” This automation, while it does not prevent spam from entering my system, enormously simplifies and expedites the process of sorting it, and making sure no genuine messages get swept out with the Trash. It therefore qualifies as a valuable supplementary solution to the problem of spam, by aiding me in its management. I imagine other e-mail clients are equipped with similar systems, which may be implemented if the user knows they are available, and takes the initiative to look into how to use them.

The down side of any automated filtering system, of course, is that with the best of “intentions” such a system may occasionally filter out messages you want to receive, but never see because for one reason or another they got flagged as “Junk.” I was out of communication for an extended period with a long-time friend, because his e-mail system had for reasons unknown flagged my e-mail sender’s address as *persona non grata* – my guess is probably because some spammer had used it as the return address for a spam distribution. The result was, I couldn’t reply to any of my friend’s repeated messages asking why I never answered his e-mails. We straightened it out eventually, after I got in touch with the people that handled his e-mail, and had the block on my e-mail address removed from his account.

The thing is, when something stands between you and the stream of information coming your way, you may never know what is being filtered out, because you never see it. On the other hand, if you don’t filter it, you’re back to standing in the flood of spam your filtering system is set up to divert. It’s an unavoidable trade-off; as with many things, “you can’t have it both ways.”

Beyond these quick and readily available at least partial solutions to the challenge of spam, there are other, more elaborate solutions that may be implemented, which involve adding some programming functionality to the site; several of which I found described among the sites that turned up in response to the Google search mentioned in **Seeking a Solution**, above. It is possible, for instance, to place one’s e-mail address exclusively on an encrypted page protected by a randomly generated password. A Perl script is put in place which is able to generate the random password on demand, and also generate on the fly a distorted graphic rendering of the same password that is legible to human eyes, but cannot be read by character recognition software. A human seeking to communicate with the owner of the e-mail address is then able to supply the matching password, enter the secure page, and use the e-mail address resident there to send an e-mail message – something an address-harvesting “spambot” is (presumably) not (yet) able to do.

There are other such solutions that may be applicable to various particular situations – and may involve concomitant trade-offs with convenience. This is not a domain in which “one size fits all.” In general, my personal bias lies on the side of simplicity, and the minimal solution that keeps things at least “within reason,” even if it may fall short of the most desirable “complete solution;” which I suspect to be a will-o’-the-wisp that will long remain beyond anybody’s actual grasp. A corollary to this bias is an inclination to put as few obstacles as possible in the way of the visitor to a Website. Therefore, if the

methods of e-mail address obfuscation described above are effective in obstructing the harvesting of e-mail addresses posted on Websites, as they seem to be, at least partially; and if the resulting (presumably greater than zero) volume of spam is within the tolerance of the owner of the e-mail address, then I am prompted to ask, "Is this good enough, or do we need to take it to the next level?" This is always the Client's choice; but in every endeavor it is possible to reach a point where "the quest for perfection" ceases to be cost-effective.

Copyright © 2008 J. Harmon Grahn. Verbatim copying and redistribution are permitted in any medium provided this notice is preserved. The HTML version of this work is available on-line at [harmonhouse.net/fdl/spam.html](http://harmonhouse.net/fdl/spam.html).