

This Free E-Book is brought to you by Natural-Aging.com.

100% Effective Natural Hormone Treatment
Menopause, Andropause And Other Hormone Imbalances
Impair Healthy Healing In People Over The Age Of 30!

Sins Of The Internet: Email Spiders

By Richard Lowe

Sins Of The Internet: Email Spiders by Richard Lowe

Warning: this article is not for the squeamish. It contains graphic descriptions of one of the biggest evils on the internet. If you can face down this evil you can reduce your load of spam by several times. Hold onto your seats and try and keep down your lunch – you are about to learn one of the secrets of how ruthless, unethical and, well, downright evil spammers steal your email address – and what you can do about it.

If you have access to your web site's log files, you will quickly find an interesting phenomenon. Your site is being visited a lot more often than you think it is. In fact, if you look closely you may be shocked to find that your HTML files are actually being used to harm you and others. In fact, you may be seeing the footprints left by some of the tools used by unscrupulous spammers to steal your email addresses.

Oh wait, let me back up a bit and explain a few things. Each time you visit a web site a record is kept of every page, graphic, sound file, video or anything else that you access (look at or download). This record is called a log file. Each line within the log file is one "hit" (other things are recorded also, but that is not important to this discussion). A "hit" is getting one "thing" from a web site. A "thing" can be an image, an HTML page, a video, a sound file or anything else. In fact, generally when you look at one HTML page you are actually "hitting" the web site many times, once for each file on the page.

Each of these lines within the log file records a number of pieces of information so that webmasters can later see what happened (don't worry, they are not generally interested in individuals – they want to know things like how many people are using Internet Explorer verses Netscape). One critical piece of information is called the "user agent". Generally this

Sins Of The Internet: Email Spiders

contains the browser name (Internet Explorer for example) or spider name (googlebot, for example, is the spider for the Google search engine).

Examine these user agent fields and you will find out many interesting facts. You will see that your site is being visited a lot more often than you would think by lots of things with strange names:

- Googlebot
- Slurp (used hundreds of search engines including Hotbot)
- Scooter (Altavista robot)
- Lycos Spider (used by the Lycos search engine)
- and many others as well.

Most of these are innocent 'bots, used by the major search engines to keep their indexes up to date. These robots are very important, for they keep your pages listed so you will get traffic. Occasionally they have other uses, including checking your pages for changes, saving your pages for offline browsing and various statistical functions.

You will also find some other names buried in your log files. These go by names such as EmailSiphon and Cherry Picker. These robots are malignant and are used by spammers to harvest email addresses. What they do is scan every single page in your web site, as fast as they can, looking for email addresses. Specifically, they are usually looking for "mailto:" type links.

Many websites have these kind of links. They are convenient, simple and create a great way for visitors to send an email to someone. In fact, it's hard to find a website which does not have email addresses embedded somewhere within the site.

In addition, people often leave their email addresses in guestbooks, message boards and other online communities which translate to web pages. Spam harvesters love these types of pages, as they can get dozens, hundreds or even thousands of different, valid and usable email addresses quickly and easily.

How do email harvesters work? Well, some scum spammer will install one of these programs on his system. He will tell it to begin scanning, which it will do rapidly and efficiently. In fact, these generally scan a web site so quickly that the server cannot do anything in the meantime (most "good" spiders, on the other hand, limit their visits to one per second, minute or even hour in order to allow other people and spiders to use the site while it is being scanned).

One of the more popular email harvester programs is called EmailSiphon (a

Sins Of The Internet: Email Spiders

product known as Sonic). The web site which promotes this garbage has the following to say:

"First of its kind on the market, Sonic helps you extract highly targeted email addresses from World Wide Web pages. Earthonline Internet marketing expertise has enabled us to program a powerful, yet sensible product that allows for proven focused lead harvesting. Therefore, Sonic with its search engine ability and single domain capability is only second in World Wide Web extraction to Earthonline Nitro."

Obviously these scumbags think they are doing a great service to the world by providing the opportunity to scan thousands of sites per day for email addresses.

Okay, so what can you do?

Ask them politely – With most "good" spiders, this is very easy to do. You

simply create a robots.txt file or use the robots metatag (if you don't know what those are, don't worry about it). Unfortunately, email harvesters are written by and used by scum, so they typically ignore polite requests.

Block them – On some web servers this is possible using special commands in a file called htaccess (again, don't worry about it if you don't know what that is), but only with those robots that clearly identify themselves. For those that don't tell you who they are (and some of them do not), then you cannot block them. In addition, the web host has to be specially set up to allow you to do this – and most, in my experience, are not.

Confuse them – Some webmasters create page after page of fake email addresses. These pages are not intended for people or good spiders (the robots metatag is used to keep the good one's out) but rather are made attractive to email harvesters. The theory is simple – the harvesters will not be able to resist the temptation (they are not very bright, as programs go) and will scan these pages. They will grab dozens, hundreds and then thousands of fake addresses, thus wasting the spammers time and possibly causing their programs to crash.

Does this work? Sure – occasionally, but it also does not prevent the spammers from getting your other email addresses, and it still chews up resources (web servers and bandwidth) sending useless messages all over the internet.

Cloak your email addresses – One thing you can do that is fairly effective is to make your email addresses look like something else. Some people create

a graphic image with the email address in it (not a great solution as it means the email address must be retyped by your visitors). Others use JavaScript to make the email address look like code. These solutions work (usually), but they make it difficult to maintain your site and often make it more difficult for your visitors. In addition, presumably the spam harvesters will eventually catch on and make their programs smarter.

Strip your site of email addresses – The only solution that works for the present time is to remove all email addresses from all of your web pages. If you need to get your visitors to send you information, then use a form (these cannot be harvested by spammers as long as the email address is not part of the form itself – Bravenet is a good service to use for this purpose). If you don't put your email addresses directly on your site, then the spammers cannot get it using their harvesters.

So there you have it. I hope this is of use to you in fighting this internet evil known as email harvesting.

Richard Lowe Jr. is the webmaster of Internet Tips And Secrets. This website includes over 1,000 free

articles to improve your internet profits, enjoyment and knowledge. Web Site Address: <http://www.internet-tips.net> Weekly newsletter: <http://www.internet-tips.net/joinlist.htm> Daily Tips: <mailto:internet-tips@GetResponse.com>

Sins of the Internet

By Richard Lowe

Sins of the Internet by Richard Lowe

One of the most common misconceptions among new (and sometimes advanced) internet users is that "if you can grab something it must be free". Add to that a touch of naiveté and a hunger for cash or fame, and you've got a formula for disaster ... or at least for some sins.

This leads directly to one of the most common and most self-destroying sins of them all – that of the Signup Junkie. Just about everyone that I know went through this phase – it's typical for most newbies. They sign up for everything. Why? Because so many companies and organizations make promises which seem to offer solutions, riches and help that it's hard to resist. The end result is, at best, lots of useless spam and at worst the loss of lots of money, time and effort.

Other sins are more subtle. For example, one day I was looking at my site statistics when I noticed something very unusual. My bandwidth usage was way too high. You see, my site had been doing about 15gb a month, which is about 500 mb/day. Suddenly my daily report showed my site was doing 2gb/day! At first I was thrilled – my site must have gotten a great ranking in the search engines or been profiled in one of the big magazines!

My joy was short lived, however, as I continued reviewing the report. I soon realized that my site was not more popular – someone had linked directly to one of my images. It was a very high traffic site so every time this page was displayed it hit my site – hammered it hard. If it kept up it became obvious that I was going to get charged for bandwidth overcharges, and it would not be a small bill.

You see, webmasters usually pay a fixed fee for their bandwidth usage up to a certain amount. In my case, it was 27gb/month. Any usage over that amount costs a small fortune, somewhere between \$6 and \$20 per gigabyte over the ceiling.

I was glad that I looked at my statistics because I was able to do something about it. I added a few lines to my htaccess file and more or less removed the link. If I hadn't looked at the reports, though, I may have had a bill of several hundred dollars.

This is an example of one of the most common and most difficult to understand sins of the internet. It's called bandwidth stealing, and it's just theft, plain and simple.

One very popular sin is that of copyright infringement. You know what I'm talking about – you've seen it all over the place. Those great Star Trek or Star Wars sites with lots of photos and scripts from the series? Guess what, they are infringing upon other's copyrights.

Spamming is very popular, and a great way to lose friends, make enemies and generally annoy people. It's not a wonderful way to make any money, as all but the truly naive hate spam and delete it immediately without reading. Spamming is a sin and has more or less lost it's usefulness long ago.

A more obscure but still popular sin is that of spamming the search engines. You see, search engines try really hard to present useful information to their users. However, some poor

misguided souls want to manipulate the system to cause their own sites to appear higher in the results list (and thus get more hits). They do all kinds of strange things to fool the little search robots, then curse when their sites are dropped forever.

If you really want to annoy your friends, you can send emails to lots of different people using the carbon copy feature. This is a common sin among newbies who don't know any better. What's the problem? Well, including a list of email addresses exposes all of those addresses to everyone receiving the message (and anyone it's been forwarded to). This often results in quite a bit of spam, as well as lots of useless messages being sent back and forth if anyone replies to the message.

Then you find people who seem to think that banners and popups are good ideas. Don't you just hate surfing to a site which consists of dozens of mindless banners and two or more popups? And worse yet, it's not like they are advertising something useful. Most often, these silly advertisements want you to purchase something that you don't need, don't want and wouldn't give to your worst enemy on a bad day.

Oftentimes, all of these sins are the result of misunderstandings as to the purpose and usefulness of the internet. I mean, you would probably not be likely to purchase furniture from a banner stapled to a telephone pole, so why would you buy one from a banner on a web site or based upon a spam email? Most people would not run power cables to their neighbors house to get their electricity, so why would they steal bandwidth?

How do you avoid all of these sins? By learning about the internet; By understanding how to use it effectively to help meet your goals; and by acting ethically and responsibly. What does that mean? Think before doing, and don't do anything you would not like others to do to you.

It's as simple as that.

Some other articles on internet sins that you may find of interest are listed below.

Sins of the internet– Bandwidth Stealing
http://www.internet-tips.net/Legal/sins_bandwidth.htm

Sins of the internet– Banners

Sins Of The Internet: Email Spiders

http://www.internet-tips.net/Legal/sins_banners.htm

Sins of the internet- Copyright Infringement

http://www.internet-tips.net/Legal/sins_copyright.htm

Sins of the internet- Not Using Bcc

http://www.internet-tips.net/Legal/sins_notusingbcc.htm

Sins of the internet- Money Making Programs

http://www.internet-tips.net/Legal/sins_moneymaking.htm

Sins of the internet- Pop Ups

http://www.internet-tips.net/Legal/sins_popups.htm

Sins of the internet- Signup Junkie

http://www.internet-tips.net/Legal/sins_signupjunkie.htm

Sins of the internet- Spamming Search Engines

http://www.internet-tips.net/Legal/sins_searchspam.htm

Sins of the internet- Webring Spammer

http://www.internet-tips.net/Legal/sins_webringspam.htm

To see a list of article available for reprint, you can send an email to:

mailto:article-list@internet-tips.net?subject=send_article_list

or visit

<http://internet-tips.net/requestarticles.htm>

Richard Lowe Jr. is the webmaster of Internet Tips And Secrets at <http://www.internet-tips.net> - Visit our website any time to read over 1,000 complete FREE articles about how to improve your internet profits, enjoyment and knowledge.



This Free E-Book has been brought to you by Natural-Aging.com.

[100% Effective Natural Hormone Treatment](#)
Menopause, Andropause And Other Hormone Imbalances
Impair Healthy Healing In People Over The Age Of 30!