

Re: WHY HAS THIS SITE BECOME SUCH A SPAM TARGET???

Source: <http://sci.tech-archive.net/Archive/sci.math/2008-03/msg03365.html>

- *From:* "Ross A. Finlayson" <raf@xxxxxxxxxxxxxxxxxxxxxxxxxxxxx>
 - *Date:* Sun, 23 Mar 2008 11:51:51 -0800
-

In doing some research through Google groups, consider something along the lines of "at3yulg2", who posted as early as Sep. 2, 2007.

In its early posts at3yulg2 promotes leadclub.com, and later atomicsearchoptimization.com. The leadclub.com seems pretty shady, consider for example how it links to "directorym.com", which has fraudulent seeming claims to wsj, inc, and other business notions. The atomicseo.net outfit advertises its "organic link-building" in terms of generating data for this search optimization type thing, i.e., "astroturf" type grassroots, fake.

X-HTTP-UserAgent: Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.0; SLCC1; .NET CLR 2.0.50727; Media Center PC 5.0; .NET CLR 3.0.04506),gzip(gfe),gzip(gfe)

The at3yulg2 first appears around September 2, 2007. Then, another interesting account is "noah harmon", along with a wide variety of other yahoo e-mail addresses that are apparently random letters. Someone has implemented a bot that post in various fora these under these various account names these various postings. Consider "sergio clive", "curtis schneider", "randolph duncan", and etcetera

<http://groups.google.com/group/Kenison-Counting-Numbers/topics?start=260&sa=N>

X-HTTP-UserAgent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322),gzip(gfe),gzip(gfe)

The first post of "noah harmon" post appears to be to "Kenison Counting Numbers", with X-IP: 222.209.126.65 , with no reverse DNS lookup (http://en.wikipedia.org/wiki/Reverse_DNS_lookup). <http://www.robert.net/ccTLD/ARPA/IN-ADDR/222> , the block's in China. (Various hosts are from many places, and many of them are probably zombie hosts, computer hosts that are virus hosts, enabling circuitous routing for track hiding, other hosts indicate the Ukraine, India, etcetera, almsot certainly innocent if ignorant dupes.)

Re: WHY HAS THIS SITE BECOME SUCH A SPAM TARGET???

<http://groups.google.com/group/Kenison-Counting-Numbers/msg/0f3374a850ae9fa7?dmode=source>

Now, all of the posts to that group except the first appear to be computer generated. The various accounts, and there aren't totally so many as there are variously many complications in setting up "free" e-mail accounts from yahoo or gmail or whoever, print copy that may have been written by a person, or could be generated by a chatbot in an easy to understand manner.

Consider these accounts names: at3yulg2, brwqvi8h, ro3z4xqh, ayt46g6b, rq5znszb, MVSLINKS, e7hrqzo, afqs64bt, rq5znszb. Now, searching for any of those reveals a list of posts by those names, and indications that they are automated spam in the response to the various usenet newsgroups where they posted. Yet, at this point, they don't have the "user profile" link next to them in the Google groups interface, where other accounts do.

So, looking for posts of September 2 2007 to sci.econ, there are a variety of them from this spam bot.

Basically I have a notion that early posts that can be attributed to this spam flood can be identified as sources, vectors, of these spam bot virii. So, then maybe various patterns towards identification of the probably very small group irresponsible for this are clearly out in the open, if an amateur like myself can find these kinds of things in a few hours.

Hi,

I wrote a program and have here lists of the spams to sci.math and sci.logic for the last week. It seems funny that there are less spams than non-spam posts through googlegroups.com, only barely. The posts through googlegroups.com are about half spam. It surprised me that in the last week around 260 posters sent through googlegroups.com, where around half of those were gmail accounts. Of the spammers 220+, pretty much all except one through googlegroups, around half of them use gmail accounts.

I wrote a program to send these posts through the abuse report web page. So, now it has sent around 900 of the spams through the web page, the last week's spams. Then, I think it would be OK for others to similarly report the abuse. The only problem is that it takes too long, it doesn't close the socket after receiving the web page. Fixing that, now that it closes the connections, it has run much faster. "Sent 859 reports."

So, now all, or almost all, of the recent (in the last week) spams through googlegroups to sci.math have been sent back to Google groups. Without some knowledge of their process, it's unclear as to whether if a particular account gets enough spam complaints, that it is suspended. If that were the case, then it would seem to be a simple matter to simply send that many.

Collecting the new spams from today, now they are being launched back to Google. There are duplicates, I haven't implemented the logic to send only new spams to Google and maintain a database of which have been sent and which haven't, the headers of the articles are just all in one big file. "Sent 905 reports." In posting the

Re: WHY HAS THIS SITE BECOME SUCH A SPAM TARGET???

Re: WHY HAS THIS SITE BECOME SUCH A SPAM TARGET???

form: http://groups.google.com/support/bin/request.py?contact_type=abuse_legal_iss&hl=en , I set the e-mail address to the spammer's address, the country of residence to USA, group name to the comma-separated list of group names (most of the Googlebot spammers post to one group at a time), the message ID to the Message-ID, type of abuse to spam and message to the headers of the spam minus path and xref.

So, I thought about arranging to send complaints on behalf of others about these spams. Again, without knowledge of Google's processes with regards to massed complaints, they know these spams are bad.

Again, there are around as many legitimate posters using googlegroups as there are spammers. There seems to be some throttling of the spam by group volume. (That does suggest a spam monoculture.)

My program is really quite inefficient at this point in partitioning spams/nonspams. Basically over the posts it examines only the posts through googlegroups.com, and then for each it has added to whitelist/blacklist by whether it is not spam or spam by inspection of the header, (actually only from and subject), and sometimes the posting patterns, i.e., the denotation as spam is not content-based. I have implemented a variety of statistics on the posts, towards implementing the supervised classifier (probably as artificial neural net, hybrid statistical learner/genetic algorithm).

Now I post the spams from the last week or so of sci.logic, "Sent 122 reports."

Ross

.