SURBL-Listed Domains And Their Registrars (In Just Seven or Eight Minutes)

Joe St Sauver, Ph.D.
(joe@uoregon.edu or joe@internet2.edu)

Forum on DNS Abuse
Grand Ballroom, 14 Mar 2011 11:30-13:00
ICANN 40, San Francisco, California

http://pages.uoregon.edu/joe/icann40/

Disclaimer: all opinions are solely those of the author, and do not necessarily represent the opinion of any other organization or entity. Data continually changes, so please do your own analysis rather than relying on the data shared by way of example in this presentation.

Some Questions of Interest

Are all ICANN-accredited registrars equally popular with spammers, or are some registrars "preferred" by spammers for whatever reason? Which registrars are being most victimized?

If only a limited number of registrars are employed for the majority of spammer registrations, convincing those few registrars to take appropriate action against spammer-controlled abusive domains could have a material direct impact on spammer operations.

Understanding spammer registrar preferences may also provide a useful metric for email reputation purposes, in the event that any given registrar is unable or unwilling to deal with problematic customers.

MAAWG SFO, February 2008

I first addressed this topic as part of a (longer) talk I gave for the 12th General Meeting of the Messaging Anti-Abuse Working Group (MAAWG), entitled "Spam, Domain Names and Registrars," see http://pages.uoregon.edu/joe/maawg12/

If you're not familiar with MAAWG, they're the leading industry anti-spam organization, with an ISP membership representing over a billion mailboxes. MAAWG's membership also includes a variety of other constituencies including legitimate senders, messaging-related product vendors, and yes, even registrars.

Participation from registrars or registries opposed to spam would be great, particularly in MAAWG's Registrar Subcommittee. A full roster of current MAAWG member organizations can be seen at http://www.maawg.org/about/roster

ObDisclaimer: I am, and have been, a senior technical advisors for MAAWG for a number of years now.

Problematic Domain Names: The SURBL List

In order to be able to map problematic domain names to registrars, we need a listing of those sort of names.

The SURBL (see http://www.surbl.org/) is an extremely well regarded project that focuses on listing domain names that have been spamvertised in the body of spam messages. Typically, SURBL will have 500,000-600,000+ domains listed at any given point in time.

If a domain name seen in an email message is listed on one or more of the SURBL zones (the SURBL has several), this is typically sufficient for Spamassassin to add from 0.122 to 4.499 points for each zone in which that domain name appears (for more information about Spamassassin tests see spamassassin.apache.org/tests_3_3_x.html). For context, at many sites, messages with an aggregate SpamAssassin score greater than 5.0 will be routinely filtered as spam.

Mapping Domain Names to Registrars

With a list of SURBL domains in hand, we randomly sorted them and then mapped them to their associated registrar. That information is typically available[0] from the registry whois. For example, looking at a SURBL-listed "MyCanadianPharmacy" site (spamtrackers.eu/wiki/index.php/My_Canadian_Pharmacy) we see:

```
% whois -n pharmacypillstablets.com
Domain Name: PHARMACYPILLSTABLETS.COM
Registrar: BIZCN.COM, INC.
[remainder snipped]
```

Note 0: Some ccTLDs may not offer whois service, or may limit whois queries to ridiculously low levels, or may have harder-to-parse whois output; in those cases, data may be unavailable for those SURBL'd domains. Other domains may already have been deleted by the time domain to registrar mapping for that domain was attempted (our mapping was done "gently" over multiple days to limit its impact).

What TLDs Did We See Most In The Feb 2011 SURBL Snapshot?

TLD	Count		Cumulative	
info	**252655**	**40.35%**	**252655**	**40.35%**
com	**187369**	**29.92%**	**440024**	**70.27%**
ru	**118646**	**18.95%**	**558670**	**89.21%**
net	32156	5.13%	590826	94.35%
cn	8488	1.36%	599314	95.70%
org	7915	1.26%	607229	96.97%

[remaining TLDs, each at less than 1.0%, are snipped]

NOTE: **Just three TLDs account for nearly 90% of all SURBL-listed domains!**

Interpretive Caution: Domains from some TLDs, while numerous in the SURBL, may only be lightly spamvertised. Domains from other TLDs, while fewer in total number in the SURBL, may be extremely aggressively spamvertised. **The counts shown above have NOT been weighted by their appearance in spam messages.**

What Registrars Did We See Most In Our Feb 2011 SURBL Snapshot?

Registrar	Count		Cumulative	
GODADDY	92055	20.49%	92055	20.49%
DOMAIN NOT FOUND	80413	17.89%	172468	38.38%
NAUNET	50755	11.29%	223223	49.68%
ENOM	42446	9.45%	265669	59.12%
REGRU	36151	8.04%	301820	67.17%
MONIKER	17191	3.83%	319011	70.99%
BIZCN.COM	6026	1.34%	325037	72.33%
NAME.COM	5945	1.32%	330982	73.66%
ONLINENIC	5341	1.19%	336323	74.84%
REGTIME	5107	1.14%	341430	75.98%
SPOT DOMAIN	5105	1.14%	346535	77.12%

[continues…]

What Registrars Did We See Most In Our Feb 2011 SURBL Snapshot? (cont.)

Registrar	Count		Cumulative	
CHINA SPRINGBOARD	5058	1.13%	351593	78.24%
DYNAMIC DOLPHIN	4398	0.98%	355991	79.22%
XIN NET	4082	0.91%	360073	80.13%
DIRECTI INTERNET	4018	0.89%	364091	81.02%
KEY-SYSTEMS GMBH	3765	0.84%	367856	81.86%
INTERNET.BS	3728	0.83%	371584	82.69%
REALTIME REGISTER	3274	0.73%	374858	83.42%
NET-CHINESE CO.	3170	0.71%	378028	84.12%
ABSYSTEMS INC	3153	0.70%	381181	84.83%
UK2 GROUP	3076	0.68%	384257	85.51%
TUCOWS	3015	0.67%	387272	86.18%

[etc.]

Should We Adjust for Registrar Market Share?

- There are some accredited registrars with a huge market share; others are quite a bit smaller. Should adjust for relative registrar size (perhaps based on www.icann.org/en/tlds/monthly-reports/)?
- We could compute (for example), the percent of all domains registered with each registrar which are listed on the SURBL. A registrar that has 65% of all of its domains listed on the SURBL would obviously have bigger challenges than one that has a fraaction of 1% of its domains listed on the SURBL.
- We could also try normalizing by market share: if a registrar had 5% of all domains that are listed on the SURBL, and also has a 5% share of the overall domain market, the ratio of those two values would be 1.0, and might be considered to be relatively "as expected." A registrar associated with 10% of domains on the SURBL, but with only 2% of the overall domain market, would have a ratio of 5.0; this might seem disproportionately high.

However…

- Scaling by size implies that there's some minimal "acceptable" or "tolerable" level of abuse. Is that true?
- If every registrar were to permit "just" 1% abusive domain names, and there are 200 million (or more) domain names worldwide, that means we'd still have at least 2 million abusive domain names, an absurdly intolerable burden for the Internet to tolerate.
- But let's be realistic: having "zero" abusive domains is probably an impossible goal to accomplish. Therefore, maybe we should be striving for something that's "small but doable," such as 0.05% abusive domains (e.g., a global total of 100,000 abusive domains).
- That would imply that Godaddy, with 37,913,616 .com, .info, .org and .net domains as of 11/2010, would have an abusive domain target (for those TLDs) of no more than 18,957 domains.
- Tucows, with 6,131,782 .com, .info, .org and .net domains, would have an abusive domain target of no more than 3,066 domains, etc.

In Conclusion, Some Possible Next Steps

- Our review was based on an un-weighted enumeration of the domains listed on the SURBL. **Scaling domain name counts by their actual observed occurrence in spam** would allow the most problematic domains to be identified (and thus the most victimized registrars) to be better identified.
- Although it is possible to patiently map domains to registrars on a domain by domain basis via whois, **it would be better if registries offered a daily domain name-to-registrar map for all domains in their TLDs**, thereby offloading public whois servers.
- Registrar domain volume information is typically NOT available for ccTLDs such as dot ru or dot cn. **More transparency with respect to registrar market share is needed for ALL TLDs.**
- We hope that the community will consider adopting a target threshold for problematic domains (such as the **0.05% threshold** mentioned), working to reduce abusive domains to that level.