

Some Ideas for the Good Domain List

mengwong@pobox.com
20040727

The goal of the Good Domain List is to make Sender ID immediately useful to a receiver. At present, RHSBLs are not very comprehensive, so a raw Sender ID installation might block some forgeries, but won't block spam sent from spammer domains. Nor will it help with the false positive problem unless a receiver already has a list of whitelisted domains. But if we augment the raw Sender ID core with a seed list of known-good domains, and include that list in the bundle, we're suddenly a lot closer to a usable Whole Product.

We can build a big list of "known good domains".

aol.com amazon.com
brightmail.com citigroup.com
earthlink.com ebay.com
gmail.com habeas.com
intel.com microsoft.com
spamhaus.org yahoo...

Good Domains

The Good Domain List can be published widely. It needs to be (1) built, and (2) maintained.

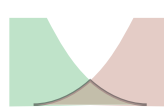
We can build a slightly smaller list of bad domains.

aaaaaaaaaspammer.com
criminalspammer.org
stock-alert-scam.com
lay-pipe-all-weekend.ru
phishing-scammer.com
419-scams-r-us.ng

Bad Domains

The Bad Domain List can be constructed internally and used to ensure that the Good Domain List is being sanely built. Publishing such a list is tricky because it would need to be updated in near-realtime. That job belongs to a proper reputation system.

Everybody else falls into the "unknown" category.



These guys are customers to the accreditation industry.

If a domain has been on an accreditation list long enough, a reputation service is likely to add it to its own whitelist.

How do we build the list?

- AOL's feedbacks
- Ironport's Senderbase
- Margaret Olson's Low Complaint Rate list
- A survey of domains that were registered prior to 1999.
- A review of a ham corpus using SPF best-guess
- Domains that have https
- Fortune 2000, etc.
- Other ideas

How do we maintain the list?

- Updates from initial feeds
- Removal of domains that have gone bad
- Other ideas

Who will build the list?

Volunteers led by Meng.

Who has volunteered to help?

Habeas.
ESPC.
Ironport.
Dennis.

Is there a business model?

The business model is a reputation clearinghouse. If the seed list of good domains turns out to be useful, there'll be demand for updates. Ultimately, a reputation service could provide a realtime snapshot of the information it gets from its reputation sources to paid subscribers, and on a time delay to the public. As a clearinghouse, the reputation service should pay its input feeds. A good input feed would report, by domain, counts of total mail vs spam complaints by users. Only messages that passed SPF (or best_guess) would be considered, of course. Such a clearinghouse should follow the Visa model and be jointly owned by its participants: the producers and the consumers both. Meng is spearheading the development of the clearinghouse model with the encouragement of other players in the reputation and accreditation space.

What might the list actually look like?

google.com	1
*.google.com	1
aol.com	1
*.ipt.aol.com	0
comcast.net	1
*.comcast.net	0

The list should ideally represent the source feeds also, perhaps obfuscated.

How will the list be offered?

The initial list will be distributed as part of SPF library bundles, and ride along any MTA distributions and packages that include SPF. Those MTAs will be configured to consult that list if the SPF result is a pass. If the sender domain is found on the list, further antispam checks can be avoided. This creates a "fast lane" for email that will support the Sender ID adoption dynamic. If a realtime provider such as Spamhaus wishes to help by offering a DNS or rsync feed of the list, it should be easy to configure the MTA to use that instead. We can also PGP-sign the list and publish it on Usenet and using P2P methods to avoid hassle.