MAAWG Briefing/Look Ahead

MAAWG 6th General Meeting
San Francisco, California, Feb 28-Mar 2nd, 2006

Joe St Sauver, Ph.D.

(joe@uoregon.edu)
MAAWG Senior Technical Advisor

http://www.uoregon.edu/~joe/maawg6/

The Background for Today's Talk

- The MAAWG senior technical advisors were all asked to reflect on "where we're currently at" and to think a little about where spam may be going in the days ahead.
- This talk is my brief take on a couple aspects of those two questions.
- While I have previously made recommendations about how carriers can scalably deal with zombied customers ("Spam Zombies and Inbound Flows to Compromised Customer Systems," http://www.uoregon.edu/~joe/zombies.pdf) that is not what we're going to focus on today. Why? There are PLENTY of other areas also worth talking about!
- Before we dive in, a brief note about the format of this talk, and a disclaimer…

Talk Format and Disclaimer

- The content of this talk has been carefully tailored for a mixed managerial/technical unvetted/public audience.
- That said, these slides are quite detailed. Why?
  --Time is limited and I find I'm prone toward getting "side tracked" and running over if I don't "stick to the script"
  -- I usually cover quite a bit of material fairly quickly
  -- I hate to be misquoted
  -- I like to provide pointers to sources for further information (but hate to make you all frantically scribble URLs)
  -- I know these slides may be viewed after the fact by those who are not here in person today, and also by those for whom English is not their primary language.
- **Disclaimer:** all opinions express in this talk are strictly my own. It would be really foolish to act on anything I suggest without doing your own "due diligence" first.

The Five Topics We'll Quickly Cover Today

1. Filtering Spam Using SURBLs
2. SMTP Auth, Port 587 and Encryption
3. Email Traffic From High Density Shared Hosting Providers
4. The Spam Filtering Complexity Threshold
5. Federal Enforcement of CAN-SPAM

I. Filtering Spam Using SURBLs

Spamvertised Domains

- Most spam (with only comparatively rare exceptions such as stock pump-and-dump spam) includes a URL (or "web page address") for a spamvertised web site which the spammer is trying to promote.
- By scanning message bodies for URLs known to be associated with spammers, it becomes possible to use the very presence of those URLs as a basis for blocking spam.
- One of the best known and most carefully administered lists of spammer URLs is the SURBL block list, see http://www.surbl.org/
- A measure of the value and trustworthiness of the SURBL data is its incorporation into the default tests done in SpamAssassin 3.1

SURBL Test Scores in SpamAssassin 3.1

- `URIBL_SC_SURBL 3.600 (SpamCop)`
  `URIBL_JP_SURBL 3.360 (Joe Wein+Prolocation)`
  `URIBL_AB_SURBL 3.306 (AbuseButler)`
  `URIBL_OB_SURBL 2.617 (Outblaze)`
  `URIBL_PH_SURBL 2.240 (Phishing)`
  `URIBL_WS_SURBL 1.533 (Bill Stearns)`
  Given that a score of 5.0 is typically sufficient for a message to be tagged or foldered as spam, clearly these are powerful tests. For comparison for those of you familiar with the traditional Spamhaus SBL and XBL DNSBLs…
  `RCVD_IN_XBL 3.114`
  `RCVD_IN_SBL 2.712`
  I will also personally tell you that the SURBL tests work REALLY, REALLY well… I suspect that spammers HATE having their URLs listed on the various SURBL sublists…

Recommendations

- **If you're not currently testing incoming mail message body URLs against the SURBL, I'd strongly urge you to begin doing so.**
- **For performance, and because of the number of queries you'll likely be making, most MAAWG-sized carriers will want to request rsync access to the SURBL. See: http://www.surbl.org/rsync-signup.html**

SMTP Auth, Port 587 and Encryption

SMTP Auth, Port 587, and Encryption

- A number of authorities, including the FTC, the Anti Spam Technical Alliance, and MAAWG itself have recommended that ISPs require their customers to submit email via port 587/TCP after the customer successfully authenticates with their username/password (e.g., as defined in RFC2476 and RFC2554), rather than just using 25/TCP. See, for example:
  -- www.ftc.gov/bcp/conline/edcams/spam/zombie/
  -- docs.yahoo.com/docs/pr/pdf/asta_soi.pdf and
  -- www.maawg.org/port25/MAAWG_Port25rec0511.pdf
- Because of the risk that traffic may be intercepted on the wire (or over wireless networks), **it is critically important that passwords NEVER be transmitted unencrypted**, whether for the purpose of SMTP Auth or for any other purpose. For SMTP Auth, this requirement is usually handled by requiring STARTTLS once an SMTP connection is made to 587/TCP.

STARTTLS vs TLS

- Unfortunately you may find that some *common* customer email clients do not correctly support STARTTLS on 587/TCP. Some mail clients may not support encryption at all; others may attempt to do SMTPS (SMTP over SSL) without doing STARTTLS negotiation (often on port 465 notwithstanding the fact that IANA has assigned 465 for use by URL Rendezvous Directory for SSM (see http://www.iana.org/assignments/port-numbers).
- Patches may help. See for example: "Outlook 2002 post-Service Pack 3 hotfix package, May 7, 2004, http://support.microsoft.com/kb/829346/EN-US/ and "Outlook 2003 post-Service Pack 1 hotfix package," August 21, 2004, http://support.microsoft.com/kb/839629/EN-US/ (Note that these are "by request only" patches, unfortunately).

Recommendations

- If you do elect to do SMTP Auth on port 587, be careful to configure all servers which will be doing SMTP Auth to require STARTTLS to avoid clear text password exposure.
- Make sure mail client software vendors know that your customers need SMTP Auth (with STARTTLS negotiation on Port 587) in their email programs.
- Document the sometimes arcane process required to enable SMTP Auth with STARTTLS for each client you recommend/support.
- Don't forget about the wireless handheld devices your customers are using, too – be sure to plan for SMTP Auth support with STARTTLS for them too!

Handling Email Traffic From High Density Shared Hosting

DNSBLs and High Density Hosting Issues

- As most of you already know, traditional DNS block lists (such as the Spamhaus SBL and XBL) list hosts by IP address, or "dotted quad."
- Unfortunately, high density "budget hosting providers" routinely serve tens or even hundreds of virtual hosting customers on a single IP address.
- Mail emitted from a mail server hosted on an IP address of that sort cannot be reliably "connected with" or "tied back to" a specific high density virtual hosting customer, thus *all* the customers on that IP address end up sharing the reputation associated with the *worst* customer using that IP address (and it is hard to avoid having at least the occasional bad customer).

Shared High Density Hosting: A Dynamic Address Pool-Like Mail Source

- This problem is very similar to the problem with dynamic address pools used by cable modem, DSL and dialup providers – there is no way to reliably accumulate reputation data for those IP addresses because a steady stream of different customers are constantly rotating through them, and as a result, many providers now routinely block mail sent direct-to-MX from known dynamic IP address ranges and from hosts with "dynamic-looking" rDNS.
- So assuming you're seeing spam from high density hosting provider IP address ranges, is it now time for you to ALSO consider blocking email sent direct-to-MX from shared high density hosting company customer IP's? (Many shared high density hosting providers currently rely on being "too big to block" and worries about "collateral damage" to avoid blocking of that sort.)

Some Hosting Company Recommendations

- Hosting companies should consider offering two types of hosting, dedicated IP hosting and shared IP hosting.
- Hosting company customers who want to emit mail direct-to-MX should be hosted on dedicated (non-shared) IP addresses. The mail server should have a fully qualified domain name consistent with its role, the forward AND reverse DNS for the server should exist and agree, and whois/rwhois entries should be created documenting each customer's IP address usage.
- Mail from shared high density virtual hosting customers should be handled just like mail from ISP dynamic address pools – the mail should be channeled through closely monitored mail servers run by the hosting company itself, and suitable anti-abuse technologies (rate limits, outbound spam scanning, SMTP auth, etc.) should be used as appropriate.

The Complexity Threshold

What's This "Complexity Threshold"?

- The spam filtering complexity threshold is the point at which average users do not, and probably cannot, realistically be expected to understand how the spam filtering that affects them actually works.
- An example of what I mean may help to illustrate this – non-technical users often assume that site-wide spam filtering must be based exclusively (or largely) on the message body "From:" email address. Of course, in reality, few if any ISP spam filters pay attention to the contents of that header because of the ease with which the value of that header can be forged…. The real mechanisms by which spam gets filtered, which are obviously far more complex than just looking at the message body "From:" headers at most sites, are unknown to most regular users and are sufficiently arcane as to be indistinguishable from magic.

It Isn't Just How Your Company Filters, Either

- Even if a user's own ISP is willing to publicly disclose the spam filtering approach they employ, and a local user takes the time to learn and understand that approach, because of the diversity of approaches used elsewhere, developing the ability to systematically diagnose and debug email delivery issues would likely require the average user to understand a large set of spam filtering methodologies in common use.
- Spam filtering has thus become a matter
  -- of **faith** ("I don't know how it works, but I trust my ISP…")
  -- of casual empirical **experimentation** ("Did this go thru?"),
  -- or a topic that requires **professional consultation** and interpretive assistance, assuming such help is available
- Hypothetically, what topics would an average user, "Uncle Bob," as it were, need to know to be able to "get" how spam is filtered today?

Some "Basic" Spam Filtering Concepts That It Would be Helpful for "Uncle Bob" To Understand

- RFC2822 (structure of email messages), and RFC2821 (the basic SMTP protocol)
- IP addresses, CIDR notation, netblocks, whois, rwhois, DNSBLs
- DNS, rDNS, static and dynamic addressing
- How content-based spam filters (such as Spam Assassin) work, scoring, false positives and false negatives
- Baeysian filtering, challenge/response and collaborative scoring approaches to spam
- Viruses, worms, phishing, backscatter, Joe jobs, 419 scams, spoofed header content, spam zombies, abuse reporting standards, accountability, reputation, whitelisting, etc., etc., etc….
- Uncle Bob just isn't going to_ bother to learn all this

What Will Uncle Bob Actually Say/Do?

- "I guess I'll just try sending a message, and see if it seems to go through or wait to see if Tom writes back. If I don't hear from him after a few days, I'll try giving him a call."
- "After I called, Tom tracked down my message to him. It had gotten stuck in his spam folder for some reason, I dunno."
- "It seems like my messages are always getting blocked. I guess I'll have to get a free web email account on Hotmail or Yahoo or Gmail and try using that instead."
- "Whoa! Email to me from my broker got blocked somehow!"
- "I'm really getting mad about all the spam I'm getting. I tried calling the ISP, but they really weren't much help. I guess I better call them again."
- "Maybe I can just buy something to fix it."
- "To heck with email – it's just too hard for me to figure out."

What About The Spammers?

- Unlike Uncle Bob, the spammers and other bad guys DO get what's going on, and they LIKE complex systems because when you have to try to routinely explain your complex systems to innocent folks, the bad guys can also get that data and use it to figure out how to most effectively do an end run around your increasingly baroque filtering.
- For example, most spammers can look at a typical rejection message and use that error to deduce what DNSBLs you're using; with that knowledge he can then pick an un-DNSBL'd host (or just keep trying one sending host after another).
- Similarly, spammers know to pre-test their spam against stock SpamAssassin, and how to adjust their draft mailing until it has a sufficiently low/deliverable score, etc. Not Bob!
- It's bad if the solution to spam befuddles the patient it is supposed to protect but doesn't affect the spammers!

Why Complexity Should Really Matter to You…

- "Complex" is a synonym for "expensive" and "fragile."
- Complex systems (for spam filtering or for other things):
  -- tend to be tricky to install, support and scale
  -- are often prone to failure
  -- can result in (expensive) customer support calls
  -- may frustrate users ==> cause customer churn.
- Complex systems also leave you vulnerable when you face management (or journalists!) who want a nice easily digested sound bite, not a bunch of technical "mumbo jumbo" they don't have the background/inclination to bother to try to understand. If you can't explain your spam filtering policy in one sentence, it's too complex.
- Of course, a trivially simple spam filtering system might not work very well in practice. (Yes, that **is** a bummer.)

Ramblings

- A major reason why third party commercial spam filtering appliances/services are so popular today is that having outsourced their spam filtering, the ISP can simply redirect all filtering-related questions to the third party for answers!
- To the extent that what you're doing is incomprehensible to laymen and/or potentially "dangerous," society's standards and expectations for your behavior and responsibilities in that sort of professional trust-based relationship tend to change… think doctors, lawyers and other professionals!
- If tips about what caused filtering to be triggered are of limited use to average users, but highly helpful to spammers, does it make sense to continue to routinely deliver that feedback to good and bad guys alike?
- Bottom line, IMHO, you really want to develop a strategy for handling the complexity of the spam filters you've deployed.

Federal Enforcement of CAN-SPAM

Federal Enforcement of CAN-SPAM

- When Congress passed CAN-SPAM, one of the things that happened was that criminal enforcement of anti-spam laws generally became a federal matter (yes, I know that there are limited statutory exceptions, but in general spam IS largely a "federal thing" at this point).
- Having spam be a "federal thing" is both good and bad.
  -- On the one hand, federal agencies are absolutely first rate/cutting edge when it comes to civil enforcement actions and criminal law enforcement – no one's better.
  -- On the other hand, federal agencies are tremendously busy and understaffed, and have many responsibilities which compete for their attention and efforts. Sometimes, I suspect -- no, I know (based on public documents as well as common sense) -- that enforcement of CAN-SPAM is simply not the FBI or the FTC's #1 priority.

Federal LE Priorities and Performance Measures

- "FBI Priorities," http://www.fbi.gov/priorities/priorities.htm does list "protecting the US against cyber-based attacks and high-technology crimes" as FBI priority number three (after only preventing terrorist attacks and counterintelligence). Wow, that's actually pretty highly ranked/pretty important….
- But what about spam per se? Does it even get mentioned in the FBI's actual Strategic Plan for 2004-2009? No… Check www.fbi.gov/publications/strategicplan/stategicplantext.htm
- Spam is also not a specific to-be-measured FBI goal with a performance target (see the FY2005 DOJ Performance and Accountability Report under Strategic Goal 2, "Enforce Federal Laws and Represent the Rights and Interests of the American People"), although some other cyber-related goals are listed (specific metrics include child porn websites closed & top ten Internet fraud targets neutralized). See www.usdoj.gov/ag/annualreports/pr2005/P2/p05-14.pdf

What Does the FTC Say About CAN-SPAM Enforcement Efforts?

- According to "Effectiveness and Enforcement of the CAN-SPAM Act: A Report to Congress," www.ftc.gov/reports/canspam05/051220canspamrpt.pdf the FTC brought 20 cases alleging violations of the CAN-SPAM act in the two years from the time when the ACT went into effect to the date that report was issued, and they mention that more than 30 additional cases were filed in federal court by the Department of Justice, state Attorneys General and Internet service providers.
- While we all sincerely appreciate each and every one of those cases, that rate of prosecution simply isn't sufficient to deal with the number of spammers who are currently active. **More investigative and prosecutorial resources need to be made available by Congress for federal CAN-SPAM enforcement activities.**

Finally, An Excellent "Must Read" Report: *U.S. Money Laundering Threat Assessment*

- Let me conclude by mentioning one new "must read" report, ***U.S. Money Laundering Threat Assessment***, http://www.dea.gov/pubs/pressrel/011106.pdf
- Once you recognize and accept that spammers are "in it for the money," understanding the money channels they're using is key to fighting them. All anti-spam folks, whether private sector or government, should carefully study the *U.S. Money Laundering Threat Assessment* and remember, "Be sure to always follow the money!"

Thanks for the Chance to Talk!

- Are there any questions?