Winning the War on Spam

Northwest Academic Computing Consortium Online Content: Policy, Strategy and Support June 5-6, 2003, Portland, Oregon

Joe St Sauver, Ph.D. (joe@oregon.uoregon.edu)
Director, User Services and
Network Applications
University of Oregon Computing Center

http://darkwing.uoregon.edu/~joe/spamwar/

I. Introduction

Our Objective:
Email "the Way It Used to Be"

- Our objective: **We want email to be "the way it used to be."** That implies that spam will be virtually non-existent, and that normal academic communication can take place with minimal games -- no need to conceal your address from most of the world, no worries about posting to mailing lists w/o munging your address, no challenge-response BS, etc….

Calibrate Your Expectations

- The war against spam won't be won overnight. It took many years for spam to get really bad; it will similarly take some time to get spam cleaned up. However, having said that, we can easily knock spam down to 1%, or 1/10th of 1% of what it would be if it were left unfiltered by following the approach outlined later in this talk. Yes, that would not be "perfect," but it may be "good enough" (at least for now).

What We'll Cover Today

- -- Just how bad ***is*** the spam problem?
  -- The decision to block (or not block) spam and the potential costs of that decision
  -- Deciding between content-based and non-content-based filtering
  -- Filtering traffic from spam-tolerant providers and from insecure hosts, and
  -- Training your users to report spam properly

Sticking To The Script

- Because we have a lot to cover, and because many spam fighting folks from your institutions who may not be attending today, I've prepared this talk in some detail and will try to "stick to the script."
- This is a good news/bad news thing: if you're looking at this presentation after the fact, you'll be able to follow what was covered; the bad news is that if you're in the audience today, there won't be a lot of "surprises" in the talk not in this handout.

II. How Bad Is The Spam Problem? Is It Really Time To Call It A "War"?

Some may be reluctant to admit they're under siege until the enemy begins flinging rotted cows over the wall with a trebuchet, but for the rest of us the signs are quite clear that yes, spammers are now waging war on us.

It's Hard To Get Good Numbers For Spam Volume...

- Something I realized while listening to parts of the recent three day FCC Spam Forum (http://www.ftc.gov/bcp/workshops/spam/) is that while spam is ubiquitous, it is hard to get good numbers concerning the total volume of spam that users see.

  There are many reasons for this...

Reasons Why Spam Volume Is Generally Poorly Quantified

- -- The percentage of email that's spam will vary from user to user, from site to site, and from day to day (so please beware of relying on statistics from one user's experience or even one site's experience)

  -- Not all sites even have the ability to distinguish spam from non-spam messages

Reasons Why Spam Volume Is Generally Poorly Quantified (2)

- -- Sites that can identify incoming email as spam will typically block email from that source (either in real time or *post hoc*).

  That blocking process will typically reduce the amount of spam seen from that source by that site (e.g., one attempted-but-blocked email connection may represent 1,000 or 10,000 or more avoided pieces of spam).

Reasons Why Spam Volume Is Generally Poorly Quantified (3)

- -- Some spam will almost always "slip through filters;" likewise some legitimate email will almost always be falsely tagged by filters as spam (this is what's usually referred to as a "false positive"). These sort of classification errors affect estimated spam volume (upwards or downwards).
- Spam volume is constantly increasing, so any spam volume estimate made today will be low tomorrow.

One Industry Estimate

- Nonetheless, one leading anti-spam firm which is able to categorize and measure spam from a large number of clients (before filtering that email for them) estimates that as of June 2003, **77.4% of all email is now spam**. See: http://www.postini.com/stats/

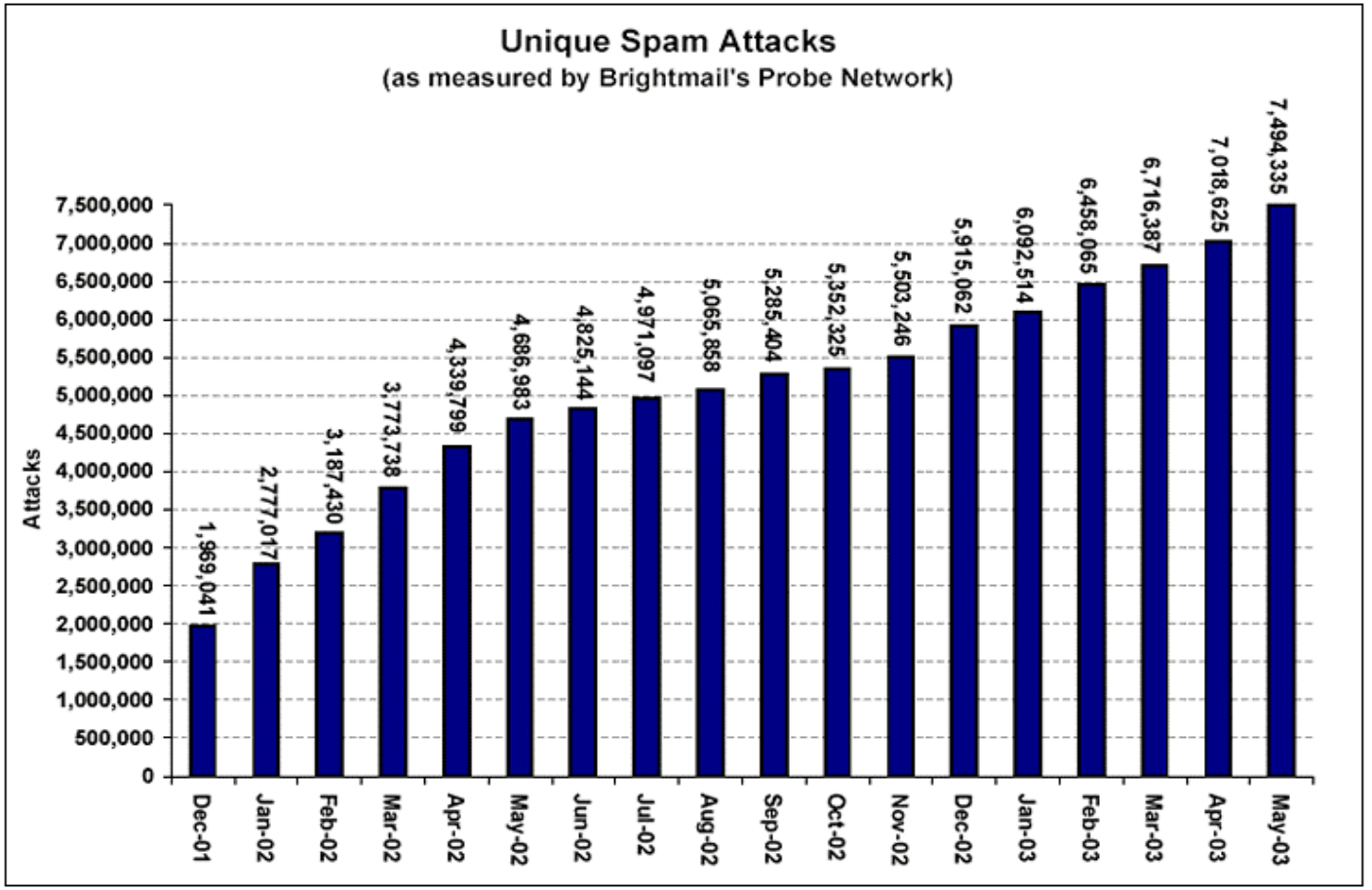
  That's three, nearly four, pieces of spam for each legitimate piece of email. That's a ***lot*** of spam.

Gartner Sez...

- "There is now 16 times as much spam on the Internet as there was just two years ago" [http://www.ecommercetimes.com/perl/story/16874.html (March 26, 2002)]
  This implies a 6 month doubling time, e.g.:
  Time=0 volume=X
  Time=6 months volume=2X
  Time=12 months volume=4X
  Time=18 months volume=8X
  Time=24 months volume=16X
  [*Time="now"* *volume=>64X?*]

Other Spam Growth Estimates

- Brightmail.com tracks what they refer to as "spam attacks." Their graph looks like:



But How Messages Are Getting Sent?

- We can talk about the percentage of mail that's spam, or spam volume doubling times, but **how many actual pieces of mail get sent on a typical day**?
- If I were to ask you to guess how much email Yahoo sends, or Hotmail, or AOL, what would you say? How many emails some of the higher volume bulk emailers send? Could you give me some numbers?

Some Top Email Senders' Message Volume Per Day

- http://www.senderbase.com/ estimates:
  -- Yahoo: 841 million/day!
  -- Hotmail: 532 million/day
  -- AOL: 331 million/day
  For comparison, some entities you've probably never heard of have traffic running:
  -- "shoppersville.net:" 288.6 million/day
  -- "ew01.com" sends 191.2 million/day
  -- "nnt6us.com" sends 188.9 million/day
- Those volumes explain a number of things...

Immense Mail Volumes ==> Limited Filtering Options

- The immense volume of real mail we're talking about on a daily basis means that for the largest of ISPs there are limited (outbound and inbound) spam filtering and abuse response options that scale to handle the volume of email they're dealing with.
- 841 million messages/day = nearly 10,000 messages a second, around the clock (except that load actually has peaks and troughs during the day)

Immense Mail Volumes ==> Some ISPs are Making Lots of Money By Hosting Spammers

- The immense volumes of bulk mail that are being sent also imply that some carriers are knowingly "bulk email friendly." There is simply no way that these ISPs could fail to know that they have customers sending huge volumes of email. ISPs knowingly host bulk email customers because those guys represent very lucrative accounts.

Spam And Exploitation Of Vulnerable Systems

- Of course, there are some spammers who are so odious that even cash-hungry carriers won't sell them service. Those guys just steal what they need, with most of them exploiting insecure systems to transmit their spam.
- For example, I'm now tracking over **300 thousand** open/abusable proxy servers, many of which are a general security risk (as well as serving as a conduit for spam).

Why Don't Those Vulnerable Hosts Get Fixed?

- Imagine that you are in charge of a large ISP's abuse desk. (You poor person!) Every morning when you come to work, there are thousands of new complaints about customers with problems -- insecure hosts that have been compromised by hackers, virus infested systems, open relays, open proxies, you name it. No matter how hard you work, more keep coming, day after day. You try to prioritize but you never catch up.

Why Don't Those Vulnerable Hosts Get Fixed? (2)

- Moreover, management tells you that you can't simply turn those users off -- these *are* paying customers we're talking about after all!! (and how much revenue comes in from those folks who are complaining about getting spammed, hmm?)
- As spam overwhelms many ISP abuse desks, a culture of ignoring **all** security problems arises; spam and other security problems seem to track very well.

So Yes, We *Are* At War...

- Spammers are ***hammering*** your mail servers ***hard*** and will ***not*** voluntarily stop.
- The problem is becoming increasingly serious (e.g., the trend line is definitely upward)
- Spam may only be the most visible symptom of many distributed security threats you are already constantly facing.
- A reasonable person would probably take steps to protect their users and systems.

III. Deciding To Act On The Spam Problem

TANSTAAFL

- The conclusion that you should take action against spam may be pretty much a matter of common sense at this point, but the decision to do so won't be without pain. (There Ain't No Such Thing As A Free Lunch).

  Understanding the tangible and intangible costs associated with the decision to fight spam will be important.

The "Collateral Damage" or "False Positive" Problem

- The most fundamental cost of blocking spam is the potential for misclassification and rejection of real, non-spam messages by anti-spam measures.
- This is normally called "collateral damage" or the "false positive" problem, and is one of the true (and unavoidable) costs of blocking spam.

When classifying mail, 4 things can happen (2 of which are bad):

- **<u>Actual Case</u>** ***<u>You Believed</u>*** **<u>The Message Was</u>**
  -- Spam Not Spam [oops! **spam got by**]
  Spam [correct classification]

  -- Not Spam Not Spam [correct classification]
  Spam [oops! **false positive**]
- We can get fewer false positives if we're more willing to let more spam slip through, <u>OR</u> less spam if we can accept more false positives. We can't minimize both objectives simultaneously

Quantifiable Costs

- Besides the pain associated with misclassifying real email (and/or failing to filter some spam messages), fighting spam will also consume direct personnel and capital resources.
- It may also come with some indirect (but potentially substantial) costs.

Costs Of Fighting Spam: People's Time

- One real cost of fighting spam is the cost of personnel, including:
  -- management time working out policies, handling complaints/inquiries, etc.
  -- systems staff time (configuring filters, etc.)
  -- end user time spent reporting spam
  -- user support/postmaster/abuse desk time
- Because nobody hires dedicated anti-spam personnel, these personnel costs are largely ignored as existing staff "just take care of it"[28]

Costs Of Fighting Spam: Hardware

- Fighting spam may entail real hardware costs:
  -- you may need to accelerate hardware upgrade schedules to insure that you have the CPU and disk space that some filtering approaches may require
  -- on the other hand, depending on how you filter, you may see a decrease in the rate at which you need to do system upgrades, as the amount of spam which gets delivered and stored goes down.

Costs Of Fighting Spam: Software/Services

- When planning the cost of fighting spam, be sure to also figure in the cost of anti-spam software products, black list subscriptions, and other software and services.
- What do all these spam-fighting costs total up to? Lacking better data, assume you'll spend the same amount fighting spam that you spend on (antivirus products plus personal firewall software) (but this is JUST a wild rule of thumb). YMMV.

Costs Of Fighting Spam: Indirect Costs

- Blocking some email may also entail some indirect institutional costs which may (ironically) dwarf direct out of pocket spam fighting costs.

Some Potential Customers May End Up Getting Blocked

- For example, if we block some "spam" directed at our admissions office, might our admissions folks miss requests for information from potential enrollees? What's the net cost to the institution if we lose tuition revenue from ten (or a hundred) potential out of state students because we're blocking their inquiry email? [Estimated UO non-resident full time tuition and fees, 2003-2004, run $16,416 per academic year.]

Blocking Spam == *Censorship?*

- While trying to block spam in good faith, you may be accused of censorship or interfering with academic freedom.
- Some approaches that may diffuse this sort of potentially explosive issue include:
  -- writing your AUP carefully to cover this
  -- allowing individual users to "opt out" from all filtering if they desire to do so
  -- delivering all email, but delivering what's believed to be spam to a different folder than presumptive non-spam email

Liability Issues?

- Are there liability issues if we don't deliver all email?
- Technical users of email used to understand that email delivery was NOT assured, and that sometimes email would NOT get through. If it did, great, if it didn't, you'd pick up the phone... Now-a-days, though, many email users seem to assume that email is an assured delivery service (even though it isn't) because it will usually get through…

Be Particularly Careful With Campus M.D.'s, Lawyers, etc.

- Under the Federal ECF (https://ecf.dcd.uscourts.gov/) email may now be used to transmit notices of legal pleadings. If email of that sort is sent to a University attorney and fails to get through, a default judgement may get entered when he misses a scheduled hearing.
- Or consider the patient of a teaching hospital surgeon who is unable to email her doc about her "chest pains," and then dies.

What If We Just Do Nothing?

- Doing nothing is equally fraught with potential problems…
  -- Spam has the potential to act as a denial of service attack against "real" mail.
  A) Real messages may easily get missed in amongst all the spam.
  B) Accounts may go over quota from spam, and begin bouncing all email.
  C) Spam sent to mailing lists you host may get sent on to subscribers, who may then unsubscribe from those lists.

Doing Nothing (2)

- -- If you choose to do nothing about spam, there will be a tremendous amount of wasted staff time as staff deal with the spam which they've been sent (plus the temptation to waste still more work time on non-work-related "content" being advertised in the spam they receive).

  What's a half hour or hour a day per employee worth to your school?

Doing Nothing (3)

- -- If you do nothing about spam, users <u>will</u> create email accounts on 3rd party services which offer ***<u>some</u>*** sort of filtering. Do you really want institutional business being done from Hotmail? Nah. Others will select and install their own spam filtering solution (good, bad or indifferent) without consulting you or anyone else.

  Doing nothing ==> email chaos reigns.

Doing Nothing (4)

- -- Keeping in mind that much spam may contain particularly objectionable sexually related content, allowing spammers unfettered access to your faculty and staff increases the chance that you may be the subject of a hostile workplace sexual harassment suit. [see: http://news.com.com/2100-1032-995658.html -- I swear I don't make this stuff up!]

Doing Nothing (5)

- If everyone else EXCEPT you filters, you are going to see a tremendous amount of spam as spammers give up on the guys who filter, and devote their attentions solely to the guys who are left.

So What Should Be Done?

- Finesse the problem. :-)
- Your best bet is probably going to be to spam filter ALL accounts by default, but allow some accounts to "opt out" and be exempt from institutionally- performed filtering on request.

Talk To Your Legal and Senior Administrative Folks

- One procedural note: whatever you decide to do about spam, be sure to talk to your university's attorneys and your senior administration folks before you implement any spam filtering strategy. Spam tends to be highly newsworthy, and there's a distinct chance you'll have a "Chronicle of Higher Education" moment if things go awry. Do NOT surprise your staff attorneys or your Chancellor/President/Provost.

III. Picking Your Defensive Anti-Spam Strategy

content-based filtering vs.
non-content-based filtering

Content Based Filtering (CBF) vs. Non-CBF filtering (NCBF)

- As you harden your systems against spam, you have a fundamental decision you need to make early on: am I going to do content-based filtering (CBF), or non-content-based filtering (NCBF)?
- Put another way, do I care about what's in the body of the message, or just about how the message arrived and possibly what's in the message headers?

One Point In Favor Of CBF...

- The biggest point in favor of CBF is that there *is* some spam which has relatively constant, readily detectable, and trivially filterable based on its content.
- If you DON'T do CBF and "easily identifiable" spam ends up getting delivered, folks will say, "How come a 'smart' computer can't ID obvious spam messages when **I** can easily do so?" This is a (sort of) legitimate complaint.

Another Advantage Of CBF

- A second advantage of doing content based filtering is that it allows you to selectively accept some content from a given traffic source, while rejecting other content from that same source.
- This can be useful if you're dealing with a large provider (such as a mailing list hosting company) that has both legitimate and spamy customers, and you don't want to end up dumping the legitimate traffic along with the spam.

CBF Issues: False Positives

- On the other hand, one of the biggest issue with CBF is the problem of false positives (mentioned previously). Because CBF uses a series of rubrics, or "rules of thumb," it is possible for those rubrics to be falsely triggered by content that "looks like" spam to the filtering rules but which actually isn't spam. For example, some (relatively crude) content based filters make it impossible for a correspondent to include certain keywords in a legitimate email message.

Using Scoring to Minimize False Positives

- Most content-based-filtering software, however, does "scoring" rather than just using a single criteria to tag spam. For example, a message in ALL CAPS might gets 0.5 points; if it also mentions millions of dollars and Nigeria, it might gets another 1.2 points; etc. Messages with a total score that exceeds a specified threshold get tagged as spam; the mere presence of a single bad keyword alone typically wouldn't be enough.

CBF Issues: The Arms Race

- Because content-based filtering attempts to exploit anomalous patterns present in the body of spam messages, there's a continuous arms race between those looking for patterns, and those attempting to avoid filtering.
- This process of chasing spam patterns and maintaining odd anti-spam heuristic rulesets is not particularly elegant, and violates the traditional desire for parsimonious solutions to scientific problems.

CBF And The Need For "Security Through Obscurity"

- For content based filtering to work well, the filters used may need to be not-well-known. For example, hypothetically, if a spammer knows that using %-encoded URLs will result in their mail getting rejected, they won't use %-encoded URLs in their spam. Thus, CBF rule efficacy may be inversely proportional to the notoriety of those rules, and preserving the effectiveness of rules may require keeping the rules used "secret."

CBF And Simplicity

- Another issue with CBF is the complexity problem: the reason why a piece of email gets filtered should be trivially comprehensible. Because many CBF's use inherently complex rules, explaining those rules to a user (or to their remote correspondent who is being filtered) can be a challenge (assuming you're allowed to disclose the basis for a given message being rejected). (Some products add X-headers with an explanation of rules that were hit)

CBF Issues: System Load

- Because CBF applies <n> unique rules to the body of each message that's received, as the number of filtering rules increases, or the size of the message body increases, or both, processing tends to slow down.
- Yes, filtering rules can be applied in order of efficacy, and arbitrary decisions can be made to limit message body examination to just the first 100KB/message (or whatever), but the scaling problem remains a real one.

CBF And The Need To Use All Rules On Each Legitimate Email

- Ironically, the more legitimate email you have, the worse CBF tends to perform.
- To understand this, note that filtering spam is a logical "OR" process -- potentially only one filter condition needs be met for a spam to be junked. On the other hand, accepting legitimate email is a logical "AND" process, and requires that a legitimate message be tested against (and successfully pass!) ALL potential filtering tests before acceptance.

CBF And Privacy

- Doing content based filtering also implicitly seems "more intrusive" to users than doing non-CBF.
- Even when CBF is done in a fully automated way, users may still be "creeped out" at the thought that their email is being "scanned" for keywords/spam patterns, etc.
- "Big Brother" is a powerful totem, whose invocation should be avoided at all costs.

Alternatives to CBF

- By now you may have the idea that I'm not a big fan of content-based filtering. You're right! So what's the alternative?
- The alternative is to focus on systems that are insecurely configured (and which are thereby unintentionally allowing spam to be sent), and providers that are bulk-email friendly (thereby intentionally allowing spam to be sent). If you block traffic from them, spam levels will drop substantially.

IV. It's Not What's In The Message, It's Where The Message Comes From That Matters

Understanding the efficiency and elegant simplicity of using non-content-based spam filtering strategies

Understanding Spammers' Fundamental Problem

- Spammers face a fundamental problem when sending spam: every piece of spam they send announces where it came from. If I know where a message came from (and I always do, at least in terms of the machine that actually handed me that message), I can put filters in place to block future mail from that source. Thus, every time a spammer uses an address to deliver a piece of spam, he puts further use of that address at risk.

Local vs. Global Blocks

- If we decide to block a site, that block could be strictly local (e.g., a file on just one system or site), or we can share that filter with a few friends, or we might submit a site that merits being blocked to one or more widely available DNS blacklists.
- Which is "better" or "worse" (from the point of view of the sysadmin of a listed system)? To be blocked system-by-system? Or to be blocked at a whole bunch of sites due to being listed in a national blacklist?

Widely Used Blacklists

- Being blocked by a widely used blacklist is quite unpleasant (from the point of view of the system administrator who's subject to that block) because suddenly a material chunk of the Internet refuses his traffic. On the other hand, being blocked via a single widely used blacklist is actually GOOD in that if he can get his system secured and delisted, his access to ALL those systems will equally suddenly be restored to normal.

Death Of A Thousand Local Cuts

- Contrast that with locally maintained filters. If a site simply drops his traffic (rather than bouncing it back to him), he may not even know that my mail is being filtered! If he's "lucky" and does learn that a site is filtering his mail, and he can persuade them to reverse that filter, that will fix that one site, but doesn't fix any of the other sites that may also be locally filtering him. Once an address becomes widely locally filtered it may never be fully usable again.

Spammers and the IP address "shortage"

- In a spammer's fantasy world (or a world where IPv6 has been widely deployed), the supply of IP addresses would be effectively limitless, and a spammer could use each address only once if he wanted to, and then move on to a new address in some unpredictable pattern to prevent anticipatory blocking. In most cases, however, it is rare for a spammer to have even a /24 (256 IP addresses) available for his direct use.

Thus, Spammers' Need for Address Gyrations

- Having comprehended that, you now understand why spammers go through such wild gyrations to launch spam at you from weird delivery channels. 1) Normal email direct from them gets blocked; 2) they can't readily move to new, as-of-yet unblocked, address space; and 3) they still desperately want to share important news about low, low, low, mortgage rates with you. They need to become "clever" to email you.

Spam Delivery Channels

- There really aren't all that many ways that spammers can send email to you:

  1) Spammers can try to send you spam via a direct connection, which will work well only until you filter those addresses. These sort of connections are the highest traffic volume spam sources you'll see, and a favorite of spammers, but comparatively easy to block for the reasons we just discussed.

Spam Delivery Channels (2)

- 2) Spammers can send you spam via a “throw away” dialup modem account. In general, direct-from-dialup spam is quite easy to block (as a category), and is now harder for spammers to use than it used to be due to ISP caller ID capabilities and credit card registration requirements. Dialups also aren't particularly good for large volume spamming due to the low speed of 56Kbps modems.

Spam Delivery Channels (3)

- 3) Spammers can send spam directly from a "throw away" web email account, such as those offered by any of hundreds of different sources. (see, for example, the lists at http://www.emailaddresses.com/)
  Most free email providers have become quite proficient at blocking large scale abuse of their service, and the ones that aren't responsible are small and generally trivial to filter. (Much spam may have a free web email "From" but not originate there)

An Aside About RFC2142

- One prime indicator of whether or not a provider is responsible is the existence of abuse@<domain>, the abuse reporting address required by RFC2142.
- If you're not sure if an ISP has an abuse address and reads complaints sent to it, see: http://www.rfc-ignorant.org/
- http://www.abuse.net lists wacky variant abuse addresses used by some providers
- Does YOUR campus have an abuse@ address? Are you on rfc-ignorant.org ?

Spam Delivery Channels (4)

- 4) Spam sent via an open SMTP relay -- this is potentially serious, but is easily identified and blocked. This was really the first serious technological thrust from spammers, and was once a far more serious problem than it is now (although it is still pops up, e.g., http://docs.info.apple.com/article.html?Artnum=106763)

  5) Spam sent via an exploitable web cgi-bin formmail script (not much of a problem, but occasionally you'll see this one pop up, too)

Spam Delivery Channels (5)

- 6) Spam sent via insecure network access points (e.g., spam sent from a lab that doesn't require authentication, open ethernet jacks or insecure wireless hubs). Potentially a very serious problem, and one that higher education is somewhat notorious for ignoring. *How many of us have live jacks with full, anonymous Internet access via unlocked classrooms, libraries, etc.? Sometimes we secure wireless, but not wired ports, etc. -- they ALL need security!*

Spam Delivery Channels (6)

- 7) Spam sent via open proxy servers (currently the most diverse and most serious spam source).

  This category includes the new "make-new-open-proxies-for-spammers-to-use" viruses such as Jeem and SoBig.

The Mechanics of Blocking

- The mechanics of blocking all these spam sources, once you understand what spam channels exist, and that there are blacklists devoted to blocking them, is relatively straightforward.

- This is the part all the geeks in the audience have been waiting for. :-)

Blocking Spamhausen

- The most convenient and elegant way to block most major spam gangs who are sending spam directly to you is through use of the SBL (Spamhaus Block List) from http://www.spamhaus.org/sbl/

  The SBL is free and can be used with sendmail or most other major mail transfer agents. Zone transfers can be arranged for large sites making 400K+ queries/day.

The SBL is NOT Spews

- A more aggressive/controversial approach to blocking known spam sources (which I don't recommend for most colleges and universities) would be to use Spews. Spews is not related to the SBL. Spews attempts to persuade spam-tolerant ISPs to be responsible by using progressively wider blacklist entries. For info on Spews, please see: http://www.spews.org/

The mail-abuse.org RBL+

- While the SBL is very good at covering what it says it will cover, it doesn't attempt to cover all the various spam channels spammers try to use. Thus, to block direct-from-dialup spam, spam sent via open SMTP relays, spam sent from some open proxies, and some additional particularly egregious spam tolerant sites, you'll want to use the mail-abuse.org RBL+ See: http://www.mail-abuse.org/rbl+/

The RBL+ Isn't Free (But It Is Cheap for .edu's)

- Not-for-profit and educational sites can license use of the RBL+ in zone transfer mode for $125/name server/year plus $5/thousand users. This translates to about $250/year for a university the size of the University of Oregon. Query mode is also available, but priced so as to discourage its use by large sites. The costs for query access is $150/name server (including 1000 users), with additional users $75 per 500.

The Particular Problem Of Open Proxy Servers

- While the RBL+ recently began to list open proxy servers, the open proxy problem is widespread enough (300,000+ known open proxy servers at this time), and so popular with spammers that it merits its own supplemental open proxy DNS blacklist. There are a number of open proxy DNSBL's available, but after considering everything, I'd recommend that you use the Easynet.nl (formerly Wirehub.nl) open proxy DNSBL.

Easynet.nl Open Proxy DNSBL

- For information about the Easynet.nl/ Wirehub.nl blackhole list, please see: http://basic.wirehub.nl/blackholes.html Note that two different Wirehub blacklists are available; the one mentioned above blocks many different spam sources; if you ONLY want to block open proxies, see: http://abuse.easynet.nl/proxies.html
- The Wirehub/Easynet list is free; zone transfers are available via rsync or wget.

Learning More About Open Proxies

- Open proxies are a fascinating topic in their own right, and one you really should learn more about (as proof of their importance and currency, they were recently discussed in the New York Times, front page, above the fold: http://www.nytimes.com/2003/05/20/technology/20SPAM.html)
- See my presentation on them this April's I2 Member Meeting in Arlington VA: http://darkwing.uoregon.edu/~joe/proxies/

Locally Maintained Filters As An Adjunct to Blacklists

- Even with all these blacklists, you may still need to augment these DNSBLs with locally maintained filters.
- If you use sendmail, you will probably implement these local filters via /etc/mail/access with entries being either problematic domains or IP address blocks. Be sure to use sendmail's delay_checks option
- Note: these files can become large. Revision control systems (like RCS) are a good idea.

Dealing With Cable Modem and DSL Customers

- While the RBL+ includes the DUL (dialup user list), which blocks mail sent directly from a dialup user's system (while allowing those users to send mail through their ISP's SMTP server), it doesn't deal with mail sent directly by cable modem and DSL customers, who should also only be sending email through their ISP's SMTP server. This is an example where locally maintained filters can really help out.

Helpful ISP DNS Conventions

- Many cable modem and DSL providers assign IP addresses to their cable modem and DSL customers that are easy to spot (such as addresses with a pattern like: <foo>.dsl.telesp.net.br). Having identified addresses of that sort it is easy to add a set of rules to /etc/mail/access which will block direct-from-DSL and direct-from-cable-modem mail. Just be sure you don't block the provider's SMTP servers!

DSL Customers and DNS

- One more caution about this -- some DSL folks seem particularly prone toward:
  1) registering a domain of their own and pointing it at their DSL connection, while
  2) failing to create a corresponding PTR (reverse DNS number-to-name) record, and
  3) failing to route their email through their provider's SMTP server.
- These guys get blocked when their dotted quad (still) resolves to <foo>.dsl.<bar>.com rather than the domain they're trying to use.

Are There DNSBLs We Shouldn't Use?

- There are DNSBLs that filter based on anything and everything; I would NOT encourage you to use every DNSBL that someone happens to offer. You really should carefully investigate the criteria used in putting hosts on and taking hosts off the DNSBL (among other things) since you're delegating a tremendous amount of authority to the operators of the DNSBLs you decide to use.

Examples of DNSBLs I Don't Use

- There are some DNSBLs that are, in my opinion, overly broad. For example, some block all traffic from Brazil, or from China. See: http://www.blackholes.us/ or http://korea.services.net/ or http://www.okean.com/asianspamblocks.html or http://countries.nerd.dk/ or http://www.cluecentral.net/rbl/).
- You are better off blocking particular vulnerabilities, or even entire spam-friendly ISPs, rather than entire countries.

V. Teaching Your Users Crucial Spam Reporting Skills

Your users are key intelligence sources in the war on spam. Your job as their leader is to train them so that they are ready to report meaningfully.

You Will Not Block ALL Spam

- No matter how good your filters, some spam will still slip through. When it does, you want to know about it so you can use that spam to help improve the blacklists you use.

  Once your users are properly trained, spammers won't be able to send spam to them without "burning" the addresses they used to do so.

Yes, You Really Do Want Your Users To Send You Their Spam

- Some of you who may already be drowning in your own personal spam may consider the idea that you want your users to send you their spam, too, to be, well, absurd. Trust me, it's not an insane idea. You NEED your users participation and cooperation because your spam may not look like THEIR spam, and besides, the sooner spam gets reported, the sooner it can get dealt with. Before long, volume will become low,

What Do Users Need To Do?

- The goal for your users is to get them to the point where they can consistently:
  -- report only spam they receive (not viruses, not legitimate message traffic), which was
  -- sent directly to one of your spam-filtered systems (not sent through some off site mailing list, departmental hosts, Hotmail, etc.),
  -- to the right local reporting address, within
  -- a day or so of the time the spam was sent,
  -- forwarded with **full/expanded headers** (and with the rest of the message body there, too).

Just Tell Us About Your Spam, Ma'am, Not Viruses

- Users sometimes have a hard time telling spam apart from virus infested messages, and may try to report both. We're really only interested in having spam reported, because (a) viruses aren't intentional, (b) we already "defang" any executable content sent via email, (c) we site license Norton Antivirus for the desktop, and (d) when we complain to ISPs about viruses, those reports seem to accomplish little or nothing.

Teaching Users To Spot Defanged Viruses

- If you defang executable attachments as we do, those executable attachments will all have a three part file name ending in .txt If you can get users to look at the file name of their attachments, you're 99% of the way there. [The other key (unrelated) part of the virus picture for them to understand is that Klez forges From: headers; have them see: http://www.wired.com/news/technology/0,1282,52174,00.html for more info.]

The Problem Of Users Reporting Legitimate Traffic

- Occasionally users will forget that they have requested email from a vendor about a particular product, or a legitimate email may have a suspicious subject line and may get reported by users wary of opening it. That sort of email obviously isn't spam, and shouldn't be reported, and for the most part doesn't tend to be, although you must be careful when rare cases do arise.

We're Sorry You're Getting Spammed on Hotmail, But...

- If your users are like ours, many of them have accounts on Yahoo or Hotmail or other 3rd party web email systems that they use in addition to their institutional accounts. That's fine, but there's nothing we can do to help with spam they get on those accounts, so please don't send it in to us. Likewise, if users are on a departmentally administered host, that's great, but again, there's nothing we can do to fix spam problems there.

Spam Arriving Via Mailing Lists

- A more common problem is spam that gets delivered to our users via some mailing list they're on that's hosted elsewhere. Assuming you use the approach we outline in this talk, spam needs to get filtered by the site hosting that mailing list; once the spam has hit a mailing list, its too late for us to do anything about it. Users need to complain to the site that's hosting the list, or convince the mailing list owner to make the list a closed or moderated list.

Using The Right Local Reporting Address

- While you may be tempted to have local users just report spam they receive to postmaster@<domain> or abuse@<domain> you really should consider creating a special spam reporting address (we'd suggest spam@<domain>) so that spam processing can be kept separate from other postmaster or abuse-related duties. (You should also avoid having users report their spam to your own personal email address.)

Getting Your Spam To Us While It Is Still Fresh

- Users need to understand that spam needs to be reported with a day or so of the time it was sent. Partially this is a matter of dealing with current issues (rather than ancient history that's already been dealt with), and partially this is a practical issue associated with some reporting services such as Spamcop (which needs you to send reports within 72 hours). Three day weekends and vacations are the biggest problem here...

Forward The Spam, Don't Use "Bounce"

- Make sure your users know to use the forward command to send you spam they receive, rather than "bouncing" it to you.
- Why? Forward preserves the integrity of the Received: headers, while bounce tends to comingle the original headers with the headers of the person bouncing the message to you, making it hard to process and report that spam appropriately.

And Then We Come To The Issue Of Full Headers...

- Anyone who works on abuse handling/spam management will tell you that the biggest obstacle to users effectively reporting their spam is getting them to enable full headers.
- My colleagues have built a nice set of how-to-enable full header pages for the email clients that our users tend to use; you're welcome to use them as the basis for local how-to-enable full header pages, too. See http://micro.uoregon.edu/fullheaders/

Providing Full Headers Is Tedious From Some Programs

- If you look through those how-to-get-full-header web pages, you'll see that getting full headers from some email products (such as MS Outlook/Outlook Express) can be very tedious, while in other cases it is a matter of pushing one button. If the email program you or your users use makes it hard to report full headers, complain to that vendor so that enabling full headers can be handled cleanly in future releases of that product.

What About End-User Training Beyond That?

- It can be tempting to just teach end users how to use Spamcop, rather than "staying in the loop" and processing the spam they report centrally. There's nothing wrong with teaching users to report spam themselves this way, but resist the urge to do it just so you don't have to see the spam that's still coming in. You need to be aware of what's still coming in, and you should feel pain if your users are getting hit with lots of spam.

End Users & Other Spam Tools

- Some users (technical enthusiasts/geeks for the most part) may also be interested in experimenting with other anti-spam tools, such as running Spamassassin.
- I would encourage you to be tolerant and encouraging toward those folks, assuming they don't become a support burden or generate system load issues (even if you wouldn't use the particular solution they're experimenting with system-wide).

What Do I Do With Spam After Users Send It In To Me?

- You may want to use http://www.spamcop.net/ to report the spam to the correct providers.
- If you subscribe to the RBL+, be sure to also report any open proxies or open relays you discover to them.
- You may want to tweak local filters
- You also can report illegal activities directly to appropriate authorities.

Oh Yes: There's Also The Issue Of User Socialization

- Beyond technical spam reporting, the other thing that you really should be doing is "socializing" your email users. By this, I mean your users need to understand:
  -- not everyone reads their email via a web browser; politeness implies that plain text (not html) is the correct way to go
  -- sending a 20MB attachment isn't something that all correspondents love getting, nor does "everyone" use Word/Excel/etc.

User Socialization (2)

- -- "Vacation" auto responders are almost always a bad idea and are seldom needed
  -- Sig files should be brief, if used at all
  -- Just because you have the technical ability (or the political clout) to send email to everyone on campus doesn't mean that you should ("intraspam" can be a real problem at some campuses)
- Helping your local users develop a culture of responsible email usage is part of getting mail back to "the way it used to be…"

The Importance Of Users Having Healthy Skepticism

- The other thing you need to inculcate in your users is a sense of healthy skepticism:
  -- No, there isn't millions of dollars waiting to be shared with you in Nigeria. Really.
  -- No, our support staff would never ask you to mail your password to a random Yahoo email account, I promise.
  -- No, the jdbgmgr.exe "teddy bear" icon isn't a virus, no matter what that chain letter from your cousin told you; do not delete it.

Grizzled Veterans Survive The Stress Of The Spam War Well

- The process of helping users become somewhat worldly and healthily skeptical is also an important component of preparing them to wage war on spam. Fighting spam can require a somewhat thick skin as you deal with disgusting message topics, and a high level of motivation as you combat an unseen and constantly morphing enemy. Skeptical/cynical "battle hardened" users are well equipped to meet those challenges.

Is There Anything We DON'T Want Our Users To Do?

- Yes. For example, we don't want them to take direct retaliatory action since they may end up mailbombing or ping flooding an innocent party who is being "Joe jobbed."
- We don't want our users to munge their address (doesn't work, can cause all sorts of support issues if done ineptly).
- We don't want users to just give up.
- We don't want users to try to intentionally solicit more spam "just for us to block." ;-)

Thanks For The Chance To Speak To You Today!

- Are there any questions?