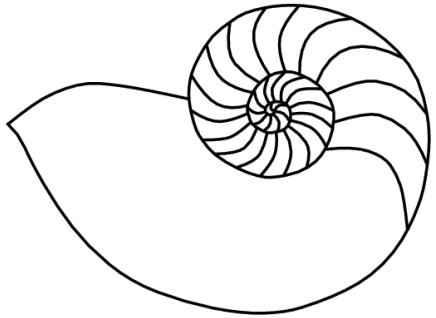
UNIVERSITY OF MANITOBA | Computer Science

Got Spam?
Fight Back!

Gilbert Detillieux

October 14, 2008
MUUG Meeting



What is Spam?

- Originally, a Usenet News phenomenon
- Name comes from
  Monty Python's Spam sketch
- Spreads to new fertile ground (e-mail, web forms, blogs, games, mobile phones, etc.)
- Currently, e-mail spam (UBE/UCE) is the biggest problem for most.

Why is it a Problem?

- Near-exponential growth rates
  - From 30 to 100 billion/day (2005-2007)
  - 85% incoming mail is “abusive e-mail” (MAAWG)
  - 90% incoming mail is spam (Spamhaus)
- Anti-spam legislation hasn’t decreased rates
  - But it might help in small ways
- Better social engineering, organized crime
  - Even knowledge workers can fall for phishing scams

What does Spam Cost?

- CPU, memory, disk space on mail servers
- Internet Bandwidth
- Lost productivity
  - $17-22 billion/year in US (2004 estimate)
  - $198 billion/year world-wide (2007)
  - $0.10/message for recipient vs $0.00001/message for sender (compare to 88% cost to sender for snail-mail)

How do Spammers Work?

- E-mail address harvesting
  - Usenet news archives
  - Web crawlers
  - Phishing, legit sign-ups, list exchanges
  - Dictionary-based & brute-force address guessing
- E-mail spam delivery
  - Free (disposable) web-mail accounts
  - Open relays
  - “Zombie” botnets

How Do We Fight Back?

- Close open relays (usually by default now)
- Content-based filtering
  - String/pattern matching
  - Statistical analysis (Bayesian filtering)
- Blacklisting and Whitelisting
- Greylisting
  (’cuz the e-mail world isn’t black & white)

How Does Greylisting Work?

- Temporarily reject unknown addresses (SMTP 400 level return codes)
- After a certain time, allow them in
- Can auto-whitelist them for return visits
- Can permanently whitelist some addresses (client, sender or recipient)
- Works because botnets typically don’t retry
- Delay may also allow them to be blacklisted elsewhere (e.g. by “honeypot” servers, etc.)

Anatomy of an SMTP Transaction

% **telnet smtp.muug.mb.ca smtp**	
Trying 130.179.31.46...	*(>TCP SYN to SMTP port)*
Connected to smtp.muug.mb.ca.	*(<SYN-ACK, then >ACK)*
Escape character is '^]'.	
220 lisa.muug.mb.ca ESMTP Sendmail 8.13.8/8.13.8; Thu, 11 Sep 2008 10:58:18 -0500	
HELO leo.muug.mb.ca	*(identify client host name)*
250 lisa.muug.mb.ca Hello lisa.muug.mb.ca [130.179.31.46], pleased to meet you	
MAIL From: <gedetil@muug.mb.ca>	*(identify sender address)*
250 2.1.0 <gedetil@muug.mb.ca>... Sender ok	*(server can accept or reject)*
RCPT To: <gedetil@muug.mb.ca>	*(identify recipient address)*
250 2.1.5 <gedetil@muug.mb.ca>... Recipient ok	*(repeat as required)*
DATA	
354 Enter mail, end with "." on a line by itself	
From: <gedetil@muug.mb.ca>	*(message headers)*
To: <gedetil@muug.mb.ca>	
Subject: test	*(add headers as required)*
↵	*(blank line)*
This is a test.	*(message body)*
.	*(dot on a line by itself)*
250 2.0.0 m8BFwIto023840 Message accepted for delivery	*(server can accept or reject)*
QUIT	*(can send another or quit)*
221 2.0.0 lisa.muug.mb.ca closing connection	
Connection closed by foreign host.	

Got Spam? My Philosophy Is…

- If mail server doesn't filter spam, humans will have to
- Software faster, more accurate
  - But… content-based methods slower, less accurate than lists
- Maximize spam rejection, but minimize false positives (getting some spam better than missing legit. e-mail)
- Server should never throw anything away
  - Reject rather than silently tossing away
  - Tag what you keep, if you think it's spam
  - Let users (or e-mail clients) worry about further filtering

My Mail Server Setup

- Red Hat-ish (RHEL, Fedora, CentOS)
- RPM packages, as much as possible
- SMTP via sendmail
  - M4 macro config, as much as possible
  - Built-in support for blacklists/whitelists
    - Local DB
    - Remote lists, via DNS
  - External mail filters (milters)

Sendmail Configuration

- Install “sendmail-cf” package
- Edit /etc/mail/sendmail.mc
  - Set up SSL support
    - Define auth. methods
    - Point to SSL key and certs
  - define(`confAUTH_OPTIONS', `A p')dnl
    - Allow relaying for authenticated connections
    - Disallow plain-text logins

Sendmail.mc Configuration *(cont.)*

- DAEMON_OPTIONS(`Port=smtp, Name=MTA')dnl
- DAEMON_OPTIONS(`Port=submission, Name=MSA, M=Ea')dnl
- DAEMON_OPTIONS(`Port=smtps, Name=TLSMTA, M=s')dnl
- FEATURE(`delay_check')dnl
  (***only*** if you'll support authenticated remote clients)
- dnl FEATURE(`accept_unresolvable_domains')dnl
  (i.e. ***disable*** this feature)

Sendmail's "access" DB

- Locally maintained
- Fast lookup
- Can blacklist or whitelist…
  - SMTP client IP addresses/ranges, domains
    - **Connect:192.168 REJECT**
    - **Connect:muug.mb.ca OK**
  - Sender addresses
    - **From:bill@microsoft.com REJECT**
  - Recipient addresses
    - **To:spamsink@honeypot.org OK**
- Problem is… ***you*** have to maintain it!

DNS Block Lists

- DNS-based blacklists of known spam senders
  - Relatively fast, cheap lookups
  - Accuracy and policies vary a lot (see stats.dnsbl.com)
- In /etc/mail/sendmail.mc:
  - FEATURE(`dnsbl', `zen.spamhaus.org', `"Open spam relay " $&{client_addr} " - see http://www.spamhaus.org/zen/"')dnl
  - FEATURE(`dnsbl', `psbl.surriel.com', `"Open spam relay " $&{client_addr} " - see http://psbl.surriel.com/"')dnl

Milter-greylist

- Install “milter-greylist” package
  - Comes with Fedora 8-10
  - Get from rpmforge/DAG for RHEL/CentOS
- Typical daemon options
  - **-P** *pidfile*
  - **-p** *socket* (usually a file-system path)
- Custom options in /etc/sysconfig/milter-greylist
  - None by default

Milter-greylist *(cont.)*

- In /etc/mail/sendmail.mc:
  - INPUT_MAIL_FILTER(`greylist', `S=local:/var/milter-greylist/milter-greylist.sock')
  - define(`confMILTER_MACROS_CONNECT', `j, {if_addr}')
  - define(`confMILTER_MACROS_HELO', `{verify}, {cert_subject}')
  - define(`confMILTER_MACROS_ENVFROM', `i, {auth_authen}')
  - define(`confMILTER_MACROS_ENVRCPT', `{greylist}')
  - define(`confINPUT_MAIL_FILTERS', `greylist')dnl

Milter-greylist *(cont.)*

- In /etc/mail/greylist.conf (order matters):
  - acl whitelist/greylist addr/domain/from/rcpt …
  - **acl whitelist addr** *my.sub.net.addr/cidr*
  - **acl whitelist from** *known-user@their.domain*
  - **acl greylist rcpt** *infrequent-user@my.domain*
  - **acl whitelist/greylist default** (do this last)
- Consider auto-generating these lists
- Can also blacklist, but why not use **access** DB?

- Multi-pronged spam filtering
  - Content-based (patterns and Bayesian)
  - DNSBL (on both headers and message body)
- Spam scoring
  - Various filters each affect overall score
  - Messages tagged as spam if score above threshold
- Client/server model (spamc/spamd)
- Typically used as a filter by procmail or e-mail clients (i.e. after e-mail received)
- Interpreted Perl code (relatively slow)

Spamass-milter +

- Install “spamass-milter” package
  - Get from Fedora EPEL for RHEL/CentOS
- In /etc/sysconfig/spamass-milter:
  - **EXTRA_FLAGS="-i 127.0.0.1,***my.sub.net.addr/cidr* **-m -r 6”**
  - **-i** says to ignore, i.e. don’t filter these nets (all your trusted subnets)
  - **-m** tells SpamAssassin not to mangle headers or message body
  - **-r 6** sets rejection threshold score to something reasonable (15 is the default)
- In /etc/sysconfig/spamassassin (no change needed):
  - **SPAMDOPTIONS="-d -c -m5 -H"**

Spamass-milter *(cont.)*

- Scores and thresholds:
  - SpamAssassin scores mail, and tags it with “X-Spam-Status:” header
  - SpamAssassin has its own threshold score (5 by default)
  - Anything above that gets tagged as spam
  - spamass-milter rejection threshold should be set higher (due to possible false-positives)
  - Anything in between is accepted for delivery, but tagged (let e-mail clients deal with it)

Spamass-milter *(cont.)*

- Spamass-milter known to crash:
  - “Mostly” stable… enough to be usable
  - Red Hat/Fedora packages come with a wrapper script to restart it
  - Other systems should use an equivalent wrapper script, or a “respawn” mechanism

Spamass-milter *(cont.)*

- In /etc/mail/sendmail.mc:
  - INPUT_MAIL_FILTER(`spamassassin', `S=unix:/var/run/spamass-milter/spamass-milter.sock, F=, T=C:15m;S:4m;R:4m;E:10m')dnl
  - define(`confMILTER_MACROS_CONNECT',`t, b, j, _, {daemon_name}, {if_name}, {if_addr}')dnl
  - define(`confMILTER_MACROS_HELO',`s, {verify}, {tls_version}, {cipher}, {cipher_bits}, {cert_subject}, {cert_issuer}')dnl
  - define(`confINPUT_MAIL_FILTERS', `greylist,spamassassin')dnl

- Content-based anti-virus scanning engine
- Uses a database of virus signatures
  - Currently, over 430,000 signatures
  - Regular updates provided
- Scans for viruses, Trojan horses & malware
  - Mostly MS Windows-based software detected
- Also scans for known phishing scams
  - Web-based, platform-neutral

- Install “clamav-milter-sysv” and “clamav-update” packages
  - Get from Fedora EPEL for RHEL/CentOS
- Comment out “Example” line in /etc/clamd.d/milter.conf and /etc/freshclam.conf
- In /etc/sysconfig/freshclam, comment out “FRESHCLAM_DELAY=disabled-warn”

ClamAV-milter *(cont.)*

- In /etc/sysconfig/clamav-milter:
  - **CLAMAV_FLAGS='-lo -P -H -k 240 \ -c /etc/clamd.d/milter.conf \ local:/var/run/clamav-milter/clamav.sock'**
  - **-lo** says scan local (from LAN) & outgoing e-mail too
  - **-P** says send warning to postmaster only
    - Users typically don't want to see these
    - If you don't either, you can use **-q** option instead
  - **-H** says include rejected-message headers in warning
  - **-k 240** says blacklist sender IP address for 4 minutes (keep this value small)

- In /etc/mail/sendmail.mc:
  - INPUT_MAIL_FILTER(`clamav', `S=local:/var/run/clamav-milter/clamav.sock, F=, T=S:4m;R:4m')dnl
  - define(`confINPUT_MAIL_FILTERS', `greylist,spamassassin,clamav')dnl

Sequence Matters…

1) Do local & DNSBL checks first…
   - … even if they seem redundant
   - fast, cheap way to do basic blocking
2) Do greylisting next
   - only moderately expensive (CPU), but memory hog
3) Do more intensive, content-based filtering last…
   - SpamAssassin is relatively slow (Perl code)
   - ClamAV also somewhat slow (and memory hog)

How Well Does It Work?

In one week…	**Count** *(on MUUG)*
Connection attempts	34,976
DNS resolution issues	134
Relaying denied	68
DNSBL rejections	19,887
Greylist rejections	14,617
SpamAssassin rejections	29
ClamAV rejections	2
Messages accepted	162

How Well Does It Work?

In one week…	**Count** *(on MUUG)*	**Count** *(on CS)*
Connection attempts	34,976	184,597
DNS resolution issues	134	1,478
Relaying denied	68	85
DNSBL rejections	19,887	109,713
Greylist rejections	14,617	46,530
SpamAssassin rejections	29	4,875
ClamAV rejections	2	141
Messages accepted	162	12,105

Questions?