Spam & Email
In Computer Science

by Hanz Makmur - makmur@cs.rutgers.edu

Laboratory for Computer Science Research

Unix-Admin Meeting • Dec 5, 2006

Available at: http://www.cs.rutgers.edu/~makmur/email-spam/
'9 out of 10 e-mails now spam'
POSTED: 5:28 a.m. EST, November 27, 2006

LONDON, England (Reuters) -- Criminal gangs using hijacked computers are behind a surge in unwanted e-mails peddling sex, drugs and stock tips.

The number of "spam" messages has tripled since June and now accounts for as many as nine out of 10 e-mails sent worldwide, according to U.S. email security company Postini.

As Christmas approaches, the daily trawl through in-boxes clogged with offers of fake Viagra, loans and sex aids is tipped to take even longer.

"Email systems are overloaded or melting down trying to keep up with all the spam," said Dan Druker, a vice president at Postini.

His company has detected 7 billion spam e-mails worldwide in November compared to 2.5 billion in June. Spam in Britain has risen by 50 percent in the last two months alone, according to internet security company SurControl.

The United States, China and Poland are the top sources of spam, data from security firm Marshal suggests.

About 200 illegal gangs are behind 80 percent of unwanted e-mails, according to Spamhaus, a body that tracks the problem.

Experts blame the rise in spam on computer programs that hijack millions of home computers to send e-mails.

These "zombie networks", also called "botnets", can link 100,000 home computers without their owners' knowledge.
“Email systems are overloaded or melting down trying to keep up with all the spam”

'9 out of 10 e-mails now spam'

POSTED: 5:28 a.m. EST, November 27, 2006

LONDON, England (Reuters) -- Criminal gangs using hijacked computers are behind a surge in unwanted e-mails peddling sex, drugs and stock tips.

The number of "spam" messages has tripled since June and now accounts for as many as nine out of 10 e-mails sent worldwide, according to U.S. email security company Postini.

"Email systems are overloaded or melting down trying to keep up with all the spam," said Dan Drucker, a vice president at Postini.

His company has detected 7 billion spam e-mails worldwide in November compared to 2.5 billion in June. Spam in Britain has risen by 50 percent in the last two months alone, according to Internet security company SurfControl.

The United States, China and Poland are the top sources of spam, data from security firm Marshal suggests.

About 200 illegal gangs are behind 80 percent of unwanted e-mails, according to Spamhaus, a body that tracks the problem.

Experts blame the rise in spam on computer programs that hijack millions of home computers to send e-mails.

These "zombie networks," also called "botnets," can link 100,000 home computers without their owners' knowledge.
The United States, China and Poland are the top sources of spam.

9 out of 10 e-mails now spam

POSTED: 5:28 a.m. EST, November 27, 2006

LONDON, England (Reuters) -- Criminal gangs using hijacked computers are behind a surge in unwanted e-mails peddling sex, drugs and stock tips.

The number of "spam" messages has tripled since June and now accounts for as many as nine out of 10 e-mails sent worldwide, according to U.S. email security company Postini.

As Christmas approaches, the daily trawl through in-boxes clogged with offers of fake Viegra, loans and sex aids is tipped to take even longer.

"E-mail systems are overloaded or melting down trying to keep up with all the spam," said Dan Druker, a vice president at Postini.

His company has detected 7 billion spam e-mails worldwide in November compared to 2.5 billion in June. Spam in Britain has risen by 50 percent in the last two months alone, according to internet security company SurfControl.

The United States, China and Poland are the top sources of spam, data from security firm Marshals suggests.

About 200 illegal gangs are behind 80 percent of unwanted e-mails, according to Spamhaus, a body that tracks the problem.

Experts blame the rise in spam on computer programs that hijack millions of home computers to send e-mails.

These "zombie networks," also called "botnets," can link 100,000 home computers without their owners' knowledge.
“About 200 illegal gangs are behind 80 percent of unwanted email.”

'9 out of 10 e-mails now spam'

LONDON, England (Reuters) -- Criminal gangs using hijacked computers are behind a surge in unwanted e-mails peddling sex, drugs and stock tips.

The number of "spam" messages has tripled since June and now accounts for as many as nine out of 10 e-mails sent worldwide, according to U.S. email security company Postini.

As Christmas approaches, the daily trawl through in-boxes clogged with offers of fake Viagra, loans and sex aids is tipped to take even longer.

"E-mail systems are overloaded or melting down trying to keep up with all the spam," said Dan Druker, a vice president at Postini.

His company has detected 7 billion spam e-mails worldwide in November compared to 2.5 billion in June. Spam in Britain has risen by 50 percent in the last two months alone, according to internet security company SurfControl.

The United States, China and Poland are the top sources of spam, data from security firm Marshall suggests.

About 200 illegal gangs are behind 80 percent of unwanted e-mails, according to SpamKiller, a body that tracks the problem.

Experts blame the rise in spam on computer programs that hijack millions of home computers to send e-mails.

These "zombie networks", also called "botnets", can link 100,000 home computers without their owners' knowledge.
"Experts blame the rise in spam on computer programs that hijack millions of home computers to send e-mails."

'9 out of 10 e-mails now spam'

POSTED: 5:28 a.m. EST, November 27, 2006

LONDON, England (Reuters) -- Criminal gangs using hijacked computers are behind a surge in unwanted e-mails peddling sex, drugs and stock tips.

The number of "spam" messages has tripled since June and now accounts for as many as nine out of 10 e-mails sent worldwide, according to U.S. email security company Postini.

As Christmas approaches, the daily trawl through in-boxes clogged with offers of fake Viagra, loans and sex aids is tipped to take even longer.

"Email systems are overloaded or melting down trying to keep up with all the spam," said Dan Drucker, a vice president at Postini.

His company has detected 7 billion spam e-mails worldwide in November compared to 2.5 billion in June. Spam in Britain has risen by 50 percent in the last two months alone, according to Internet security company SurfControl.

The United States, China and Poland are the top sources of spam, data from security firmMarshal suggests.

About 200 illegal gangs are behind 80 percent of unwanted e-mails, according to Spamhaus, a body that tracks the problem.

Experts blame the rise in spam on computer programs that hijack millions of home computers to send e-mails.

These "zombie networks," also called "botnets," can link 100,000 home computers without their owners' knowledge.

ADVERTISER LINKS

CRM Suite Software at Maysoft.com
Maysoft.com offers anti-spam software. Try our software on Domino Server Version...
www.maysoft.com

Astaro Security Gateway
All-In-One Security Appliances VPN, E-Mail and Web Ant-Virus, Spyware,...
www.m5systems.net

Mailenable Mail Servers for MS Windows
Unlimited domains and users. Includes
Observation

- Mail to Unknown Users ↑
- Over quota errors ↑
- Mail queues ↑
- Bounces ↑
- Spam ↑
- Slow Delivery
RU DCS Mail Statistics

**Total of Incoming Mail in 24 Hours**
- Bad Mail (Avg: 53337 Max: 1138428 Min: 19954 messages)
- Good Mail (Avg: 9464 Max: 24557 messages)

**24 Hours Ratio of Incoming Mail**
- Bad (Avg: 83 % Max: 99 % Min: 71 %)
- Good (Avg: 17 % Max: 29 % Min: 1 %)
Problem 1: Infrastructure

- Multiple mail servers
  - Cluster of servers
    - Faculty, research, grad, undergrad etc.
    - Each cluster server accept email

- Clients MX point to cluster server
  - Mail to client will be received by server
Problem 2: Too many addresses

- Multiple email addresses
  - username@{client1…clientx}
  - username@{cluster1…cluster2}

- Multiple Entry points
- Unread email
- Waste of resources
Problem 3: Overloading

- Overloaded Mail Queues
  - Slow delivery

- Overloaded spam servers
  - Email gets reprocessed
  - High volume → high load
  - Spam servers were timing out
Problem 4: Interface

- Primitive interface
  - Text only, no webmail
- Unreliable IMAP protocol
  - File locking problem
  - Restricted disk space
  - Reappearing mail
Flow of RU DCS Email

Incoming mails

Client1

Clusters

spamfilter

mail server

User mailboxes

Before Aug 1, 2006
What To Do?

- Email overhaul
  - Consolidate all mail servers to one
- Better spam processing
  - Reduce timeout
  - Reduce False Positives/Negatives
- Improve quality of service
  - Faster delivery, better spam rejection/detection
  - Reliable and modern Interface
- Announced May 1, deployed Aug 1, 2006
Email Changes

- Single mail server
  - Communigate Pro
- New email policy
- Self Account Creation
- Min. 1 GB. Highest 6 GB.
  - warning at 80%, 95% quota
- Centralized spam detection
  - One time check
Spamfilter Setup

- Server based SpamAssassin
- 3 Servers accessed via spamc
- DNS round robin
SpamAssassin Enhancements

- **Dynamic Custom Rules Set**
  - RulesEmporium.com’s rules
  - Updated daily like Virus DAT files

- **Optional Modules**
  - Collaborative network: DCC, Razor2
  - Signature detection: DomainKey, SPF
  - Custom Plugins: IPCountry, fuzzyOCR*

*added Oct 2006
Delay SMTP prompt
Immediate rejection if in spamhaus
Automatic blacklist on failures
No repeat spam checking
No auto replies for lists or bulk mail
Banned headers: Message*$6c822ecf@*
Other Enhancements

- Localized DNS Block List checks
- Caching nameserver
- Immediate rejection to non server
New Flow of RU DCS Email

Incoming mails

client

Nomail

Mail server

User mailboxes

spamfilter

After Aug 1, 2006
Tagging Spam Email

Little outburst, but why will you be sending money to that man. It are respected; I will make the same my business, as I have all through. him, I think, the worst of all. And it used to cut me to the quick to presently she broke out, And what is the meaning of all this? Why is

YOU'VE SEEN IT BEFORE YOU SAY?
YOU AIN'T SEEN NOTHING!
THIS IS AN ALERT FOR MONDAY NOVEMBER 13!

Company Name: MOTION PICTURE GROUP ("MPRG.PK")

Symbol: MPRG
Price: $0.20
Target: $1.0

Current Recommendation: Very Agressive Buy

MPRG RELEASES BREAKING NEWS:

The Motion Picture Group (MPRG) has released a number of new films including the Feature Film/Documentary "I Trust You To Kill Me" Staring Keifer Sutherland. The World Renowned William Morris Agency, represented the sale and yesterday they announced huge success at the AFM.

This one is selling at $0.20 but wont stay that way for long, we are expecting the details of the sale at the open of next week. Waste no time, Keifer Sutherland is one of the hottest actors in Hollywood; this sale could push the stock through the roof. Grab it first thing Monday morning!

MPRG IS THE ULTIMATE STOCK IN THE MOVIE INDUSTRY!
WATCH MPRG ON NOV 13!
Tagged Mail Headers

X-Spam-Flag: No
X-Spam-Checker-Version: SpamAssassin 3.1.6 (2006-10-03) on spamfilter2.rutgers.edu
X-Spam-Level: x
X-Spam-Status: No, score=1.1 required=5.0 tests=BAYES_40,EXTRA_MPART_TYPE,
  autolearn=disabled version=3.1.6
X-Spam-Report:
  * 1.1 EXTRA_MPART_TYPE Header has extraneous Content-type:...type= entry
  * 0.0 BAYES_40 BODY: Bayesian spam probability is 40 to 60% [score: 0.5309]

When checked during an Internet network problem
Tagged Mail Headers

X-Spam-Flag: No
X-Spam-Checker-Version: SpamAssassin 3.1.6 (2006-10-03) on spamfilter2.rutgers.edu
X-Spam-Level: xxx
X-Spam-Status: No, score=3.1 required=5.0 tests=BAYES_40,EXTRA_MPART_TYPE,
               RCVD_IN_NJABL_DUL,RCVD_IN_SORBS_DUL,
               autolearn=disabled version=3.1.6
X-Spam-Report:
   * 1.1 EXTRA_MPART_TYPE Header has extraneous Content-type:...type= entry
   * 0.0 BAYES_40 BODY: Bayesian spam probability is 40 to 60% [score: 0.5309]
   * 1.0 RCVD_IN_SORBS_DUL RBL: SORBS: sent directly from dynamic IP address
     [85.100.150.234 listed in sorbs.cs.rutgers.edu]
   * 1.0 RCVD_IN_NJABL_DUL RBL: NJABL: dialup sender did non-local SMTP
     [85.100.150.234 listed in njabl.cs.rutgers.edu]

When checked during an Internet network problem
and local DNS Block Lists
Tagged Mail Headers

X-Spam-Flag: YES
X-Spam-Checker-Version: SpamAssassin 3.1.6 (2006-10-03) on spamfilter2.rutgers.edu
X-Spam-Level: xxxxx
X-Spam-Status: Yes, score=5.1 required=5.0 tests=BAYES_40,EXTRA_MPART_TYPE,
RELAYCOUNTRY_CN, RCVD_IN_NJABL_DUL, RCVD_IN_SORBS_DUL,
autolearn=disabled version=3.1.6
X-Spam-Report:
* 1.1 EXTRA_MPART_TYPE Header has extraneous Content-type:...type= entry
* 0.0 BAYES_40 BODY: Bayesian spam probability is 40 to 60% [score: 0.5309]
* 1.0 RCVD_IN_SORBS_DUL RBL: SORBS: sent directly from dynamic IP address
  [85.100.150.234 listed in sorbs.cs.rutgers.edu]
* 1.0 RCVD_IN_NJABL_DUL RBL: NJABL: dialup sender did non-local SMTP
  [85.100.150.234 listed in njabl.cs.rutgers.edu]
* 2.0 RELAYCOUNTRY_CN Relayed through China

When checked with DNS Block Lists and IPCountry Plugin.
Tagged Mail Headers

X-Spam-Flag: YES
X-Spam-Checker-Version: SpamAssassin 3.1.6 (2006-10-03) on spamfilter2.rutgers.edu
X-Spam-Level: xxxxxx
X-Spam-Status: Yes, score=6.6 required=5.0 tests=BAYES_99,EXTRA_MPART_TYPE,
MY_CID_AND_STYLE,RELAYCOUNTRY_CN, SARE_GIF_ATTACH
RCVD_IN_NJABL_DUL, RCVD_IN_SORBS_DUL,
autolearn=disabled version=3.1.6
X-Spam-Report:
  * 1.1 EXTRA_MPART_TYPE Header has extraneous Content-type:...type= entry
  * 0.0 BAYES_99 BODY: Bayesian spam probability is 40 to 60% [score: 0.5309]
  * 1.0 RCVD_IN_SORBS_DUL RBL: SORBS: sent directly from dynamic IP address
    [85.100.150.234 listed in sorbs.cs.rutgers.edu]
  * 1.0 RCVD_IN_NJABL_DUL RBL: NJABL: dialup sender did non-local SMTP
    [85.100.150.234 listed in njabl.cs.rutgers.edu]
  * 2.0 RELAYCOUNTRY_CN Relayed through China
  * 0.8 SARE_GIF_ATTACH FULL: Email has a inline gif
  * 0.7 MY_CID_AND_STYLE SARE cid and style

When checked with DNS Block Lists, IPCountry Plugin
and additional rules from Rules Emporium
There you are, young Copperfield, and a royal spread you've got. confused. If I had any doubt of him, I suppose this half awakened was treated well here, I should beg acceptance of a trifle, instead Copperfield, to be left till called for? said the guard. Come.

TRADING ALERT!
WE SEE A RUN STARTING TO HAPPEN.
IT'S JUST BEGINNING OF 1 WEEK PROMOTION!

Trade Date: Monday, November 20, 2006
Company: PRG Group Inc.
Symbol: PRGJ
Current Price: $1.33 UP $0.19(17%) on Friday
3-Day Target: ! $5 !
Rating: 10/10
Recommendation: STRONG BUY

We want to congratulate all of our members who read our PRGJ Trading Alert on Friday. We think the fun is just beginning with this stock. Do not wait until it is too late!!! Remember this is a STRONG BUY RECOMMENDATION...
Sample Headers

X-Spam-Flag: No
X-Spam-Checker-Version: SpamAssassin 3.1.6 (2006-10-03) on sspamfilter2.rutgers.edu
X-Spam-Level: xxxx
X-Spam-Status: No, score=4.8 required=5.0 tests=BAYES_50,EXTRA_MPART_TYPE,
   HTML_30_40,HTML_MESSAGE,MY_CID_AND_ARIAL2,
   MY_CID_AND_STYLE, MY_CID_ARIAL_STYLE, SARE_GIF_ATTACH
autolearn=disabled version=3.1.6
X-Spam-Report:
   * 1.1 EXTRA_MPART_TYPE Header has extraneous Content-type:...type= entry
   * 0.4 HTML_30_40 BODY: Message is 30% to 40% HTML
   * 0.0 HTML_MESSAGE BODY: HTML included in message
   * 0.0 BAYES_50 BODY: Bayesian spam probability is 40 to 60% [score: 0.5562]
   * 0.8 SARE_GIF_ATTACH FULL: Email has a inline gif
   * 0.7 MY_CID_AND_STYLE SARE cid and style
   * 1.1 MY_CID_ARIAL_STYLE SARE cid arial2 style
   * 0.7 MY_CID_AND_ARIAL2 SARE CID and Arial2
Sample Headers With OCR

X-Spam-Flag: YES
X-Spam-Checker-Version: SpamAssassin 3.1.6 (2006-10-03) on sspamfilter2.rutgers.edu
X-Spam-Level: xxxxxxxxxx
X-Spam-Status: Yes, score=9.7 required=5.0 tests=BAYES_50,EXTRA_MPART_TYPE,
    FUZZY_OCR_KNOWN_HASH,HTML_30_40,HTML_MESSAGE,MY_CID_AND_ARIAL2,
    MY_CID_AND_STYLE, MY_CID_ARIAL_STYLE,SARE_GIF_ATTACH
autolearn=disabled version=3.1.6

X-Spam-Report:
  * 1.1 EXTRA_MPART_TYPE Header has extraneous Content-type:...type= entry
  * 0.4 HTML_30_40 BODY: Message is 30% to 40% HTML
  * 0.0 HTML_MESSAGE BODY: HTML included in message
  * 0.0 BAYES_50 BODY: Bayesian spam probability is 40 to 60% [score: 0.5562]
  * 0.8 SARE_GIF_ATTACH FULL: Email has a inline gif
  * 0.7 MY_CID_AND_STYLE SARE cid and style
  * 1.1 MY_CID_ARIAL_STYLE SARE cid Arial2 style
  * 0.7 MY_CID_AND_ARIAL2 SARE CID and Arial2
  * 4.9 FUZZY_OCR_KNOWN_HASH BODY: Mail contains an image with known hash
     * Words found:
       * "buy" in 1 lines
       * "strongbuy" in 1 lines
       * "rating" in 1 lines
       * "november" in 1 lines
       * "price" in 1 lines
       * "alert" in 1 lines
       * "thefunisjust" in 1 lines
       * (7 word occurrences found)
Evolving Image Spam
X-Spam-Flag: YES
X-Spam-Checker-Version: SpamAssassin 3.1.6 (2006-10-03) on spamfilter2.rutgers.edu
X-Spam-Level: xxxxxxxxxxxxxxx
X-Spam-Status: Yes, score=16.7 required=5.0 tests=BAYES_60,EXTRA_MPART_TYPE,
                 HTML_30_40,HTML_IMAGE_ONLY_28,HTML_MESSAGE,MY_CID_AND_ARIAL2,
                 MY_CID_AND_CLOSING,MY_CID_AND_STYLE,MY_CID_ARIAL2_CLOSING,
                 MY_CID_ARIAL_STYLE,RCVD_IN_BL_SPAMCOP_NET,RCVD_IN_NJABL_DUL,
                 SARE_GIF_ATTACH,SARE_GIF_STOX,TW_JS autolearn=disabled version=3.1.6
X-Spam-Report:
* 1.1 EXTRA_MPART_TYPE Header has extraneous Content-type:...type= entry
* 0.1 TW_JS BODY: Odd Letter Triples with JS
* 0.4 HTML_30_40 BODY: Message is 30% to 40% HTML
* 1.9 HTML_IMAGE_ONLY_28 BODY: HTML: images with 2400-2800 bytes of words
* 1.0 BAYES_60 BODY: Bayesian spam probability is 60 to 80%
* [score: 0.6659]
* 0.0 HTML_MESSAGE BODY: HTML included in message
* 0.8 SARE_GIF_ATTACH FULL: Email has an inline gif
* 4.3 RCVD_IN_BL_SPAMCOP_NET RBL: Received via a relay in bl.spamcop.net
  * [Blocked - see <http://www.spamcop.net/bl.shtml?81.103.37.107>]
* 1.0 RCVD_IN_NJABL_DUL RBL: NJABL: dialup sender did non-local SMTP
  * [81.103.37.107 listed in njabl.cs.rutgers.edu]
* 0.9 MY_CID_AND_CLOSING SARE cid and closing
* 0.7 MY_CID_AND_STYLE SARE cid and style
* 1.2 MY_CID_ARIAL2_CLOSING SARE cid arial2 closing
* 1.1 MY_CID_ARIAL_STYLE SARE cid arial2 style
* 0.7 MY_CID_AND_ARIAL2 SARE CID and Arial2
* 1.7 SARE_GIF_STOX Inline Gif with little HTML
Current State

- Processed spam ↓, actual spam ↑
- More efficient use of resources
- Faster & better mail service
  - Webmail, reliable imap, large quota

![Graph showing details of incoming mail (1 Day average)]

- %Ham ↓
- %Spam ↓
- %Rejection ↑

Processed >70%
Processed <35%
Daily RU DCS Mail Stats

Details of Incoming Mail (1 Hour average)

- Bad SMTP commands
- Processed
- Unknown users

Details of Incoming Mail (1 Day average)

- Attacks stopped by Spamhaus DNSBL
- Volume ↑ 4x

- RejectedFrom: 8285
- RejectedCmd: 12321
- RejectedTo: 61564
- Rejectedbody: 798
- IncomingSpam: 23288
- Virus: 193
- Clean: 8141
- Bad: 106438
- Total: 114583

- RejectedFrom: 9153
- RejectedCmd: 5033
- RejectedTo: 27876
- Rejectedbody: 1205
- IncomingSpam: 22907
- Virus: 467
- Clean: 9203
- Bad: 66641
- Total: 76194

Volume ↑ 4x
Review

- **Problems**
  - Multiple Servers and Entry Points
  - Overloaded Servers
  - Inadequate quality of mail service

- **Solutions:**
  - Single server and single entry point
  - Better spam processing
  - Improve quality of mail service
Future

- Problem is not going away
  - Arms race
- SMTP assume trustworthiness
  - The net is not secure
- Better solution needed
Questions?

Notes:
- **This presentation**
- **RU DCS Mail stats**
  - [http://report.rutgers.edu/mrtg/mail/](http://report.rutgers.edu/mrtg/mail/)
- **DCS Spamfilter**
- **SpamAssassin modules/plugins**
- **RulesEmporium - additional rules**
  - [http://rulesemporium.com/](http://rulesemporium.com/)
- **CommunigatePro mail server**
What is DNSBL?

- DNS BLOCKLIST
- Simple zone file
  - 12.107.239.0/24
  - 12.107.246.0/23
  - adult-news.biz
  - adultzone-xxx.com
Available DNSBL Zones

- SURBL.cs.rutgers.edu
- SORBS.cs.rutgers.edu
- NJABL.cs.rutgers.edu
- URIBL.cs.rutgers.edu
- SBL-XBL.cs.rutgers.edu
- DSBL.cs.rutgers.edu
- COMPLETEWHOIS.cs.rutgers.edu
Other SpamAssassin DNSBL

- mail-abuse.org
- sattrusted.bondedsender.org
- iadb.isipp.com
- sa-accredit.habeas.com
Running Your Own DNSBL

- Fast, Cheap and Easy
- Setup and Forget
- Reliability
What is Needed?

- RblDNSd & Rsync
- A Linux machine
- Donation to RBLDNS fund
Setting Up rbldnsd

- Install rbldnsd
- Rsync zone files
- Setup Cronjob
- Add entries to RU DNS
- Set new rbldns in spamassassin
Questions

- **URLS**
  - [http://please.rutgers.edu/show/dnsbl/](http://please.rutgers.edu/show/dnsbl/)