



## Fighting Spam

**Derek Rogillio**  
**Senior Consultant**  
**derek (at) netcraftsmen (dot) net**

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## About the Presenter

- Derek Rogillio
  - Senior Consultant
  - Chesapeake NetCraftsmen, LLC
  - CCIE #6146, JNCIE #42
  - derek (at) netcraftsmen (dot) net

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Agenda

- Introduction to Unsolicited Bulk Electronic Mail (UBE)
- Stopping Spam
- Tracking, Blocking, and Filtering Spam
- Spam Filtering Architectures and Examples

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Introduction to Unsolicited Bulk Electronic Mail (UBE)

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## What is UBE?

- Well, spam of course!
- From dictionary.com:

spam (n): Unsolicited e-mail, often of a commercial nature, sent indiscriminately to multiple mailing lists, individuals, or newsgroups; junk e-mail.

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Why All the Fuss?

- Additional hassle and loss of productivity for individuals and employees
- Example:
  - 15,000 employees
  - 50 spam messages a day
  - 2 seconds to review and delete
  - **More than 2000 hours of productivity lost per week!**

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Inappropriate Content

- Often contains inappropriate content for the workplace
  - Examples on the following slides ...

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Extra E-Mail Goodies

- Sometimes used to carry virus, trojan, or spyware software
- Do you use an e-mail application that supports auto-preview?

```

```

**Web Bug!**

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Mail Server Load

- Additional load on Internet mail servers
  - Network bandwidth
  - Processing overhead
  - Disk space

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Who Gets the Bill?

- One of the few types of commercial advertisement where the receiver bears most of the cost

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Minimum Requirements for Becoming a Spammer

1. Customer who would like to advertise via unsolicited e-mail
2. Spam software application
3. List of recipients (**victims!**)
4. Internet access

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Profile: Alan Ralsky, Spam King

- One of the most successful and prolific spammers
- Lives in West Bloomfield, MI in his new \$740,000 house
  - 8,000 square feet

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Customers

- Mr. Ralsky refuses to send UBE promoting pornography or sexual messages
- Primary customers consist of on-line casinos, mortgage refinancing services, vacation promotions, and Internet pharmacies

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Software Application

- Custom application designed by his system administrator, Charlie Brown
- Run from more than 20 computers in his home

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Mailing List

- Consists of more than 250 million mail addresses
- A full-time employee maintains the list
  
- Charge to send one solicitation to his entire list: up to \$22,000

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Internet Access

- Controls ~190 e-mail servers:
  - ~110 in Southfield, MI
  - ~50 in Dallas, TX
  - ~30 in Canada, Russia, China, and India
- Routed primarily through overseas Internet providers
- Capable of sending 650,000 messages per hour – over a billion per day

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Quotes

- "I'll never quit. I like what I do. This is the greatest business in the world."
- "There is no way this can be stopped. It's a perfectly legal business that has allowed anybody to compete with the Fortune 500 companies."
- "When you're sending out 250 million e-mails, even a blind squirrel will find a nut."

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## A Sobering Thought

- Mr. Ralsky has amassed his fortune with an e-mail response rate of less than one quarter of one percent
- During the time he has been in business, spam has increased from 8% to 36% of all electronic mail. It is expected to increase to 50% by 2005.

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Stopping Spam

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## What Can We Do?

1. User education!
  2. Make it illegal to send spam
  3. Make it technically impossible to send spam
  4. Make it technically impossible to receive spam
- Hit the spammers where it hurts the most!

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Option 1: User Education

- If no one buys, the spammers are out of business!
- Examples of important training:
  - “Remove me from this list”
  - Dangers of auto-preview and HTML mail
  - How to keep their e-mail addresses private
- Center for Democracy and Technology
  - <http://www.cdt.org>
  - *Why Am I Getting All This Spam?*

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Option 2: The Legal Avenue

- Very good resource at <http://www.spamlaws.com>
- In the US, it could be decided by case law rather than by legislative bodies
  - Five suits filed by AOL on 4/15 are a good example!
- Legislating spam away only affects the spammers who care about the law

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Option 3: Stopping Spammers From Sending Spam

- Simple Mail Transfer Protocol (SMTP) is used to transfer e-mail across the Internet
- Designed when the Internet was small and friendly
- Some extensions have been made to help, but a redesign is necessary to block everything

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Role of the Internet Service Provider

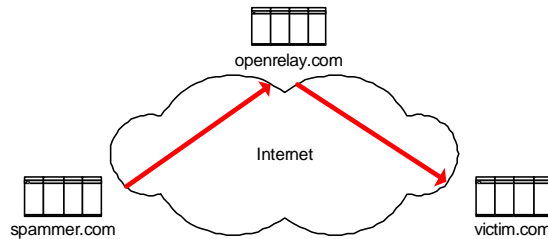
- ISPs are currently the most important entities that can stop spam on the front lines
  - Expensive to implement
  - Difficult to maintain
  - Often inconvenient for users

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Open Mail Relays

- Open mail relays and proxies are a serious impediment to stopping spam!



© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Blocking “Invalid” Mail Servers

- Becoming a popular method for blocking spam
- Mail servers configured to block:
  - Mail from DSL, cable, dialup, and other dynamic addresses
  - Mail from servers whose reverse DNS lookup doesn't match their offered domain

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Option 4: Blocking the Receipt of Spam

- The best method available today for preventing users from being overwhelmed with spam
- Server receives the mail as usual
- Filters it with a variety of techniques before delivery to the end user
- We will cover some of the strategies for the remainder of this session

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Tracking, Blocking, and Filtering Spam

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: Standard E-Mail Message

Date: Wed, 15 Jan 2003 12:00:48 -0500 (EST)  
From: John Smith <jsmith@server.rogillio.net>  
To: Derek Rogillio <derek@server.rogillio.net>  
Subject: Not Spam

This e-mail message is legitimate!

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: All Message Headers

From jsmith@server.rogillio.net Wed Jan 15 12:00:50 2003  
Return-Path: jsmith Received: (from jsmith@localhost) by  
server.rogillio.net (8.6.12/8.6.9) id MAA00135; Wed, 15 Jan  
2003 12:00:48 -0500  
Date: Wed, 15 Jan 2003 12:00:48 -0500 (EST)  
From: John Smith <jsmith@server.rogillio.net>  
To: Derek Rogillio <derek@server.rogillio.net>  
Subject: Not Spam  
Message-ID: <Pine.LNX.3.91.970325115954.130A-  
100000@server.rogillio.net>  
MIME-Version: 1.0  
Content-Type: TEXT/PLAIN; charset=US-ASCII  
Status: RO  
X-Status:

This e-mail message is legitimate!

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: Forged Mail Headers

Date: Tue, 25 Mar 1997 12:25:57 -0500

From: nobody@nowhere.net

Could I find a forged message that is more obvious?

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Forged Mail Headers: Who Is Responsible?

From nobody@nowhere.net Wed Jan 15 12:26:29 2003

Return-Path: nobody@nowhere.net

Received: from nowhere.com (jsmith@localhost [127.0.0.1]) by server.rogillio.net (8.6.12/8.6.9) with SMTP id MAA00153 for derek; Wed, 15 Jan 2003 12:25:57 -0500

Date: Wed, 15 Jan 2003 12:25:57 -0500

From: nobody@nowhere.net

Message-Id: <199703251725.MAA00153@server.rogillio.net>

Apparently-To: derek@server.rogillio.net

Status: RO

X-Status:

Could I find a forged message that is more obvious?

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: Actual Spam

```
Return-Path: <tammyhardyfjbb@mindless.com>
Received: from mindless.com ([202.7.209.122]) by
server.rogillio.net (rogillio.net mail service) with SMTP id
18xLfy7t43Nl3oW0 Sun, 12 Jan 2003 11:46:12 -0500 (EST)
Received: from 115.131.120.61 ([115.131.120.61]) by
webmail.halftomorrow.com with esmtp; Sun, 12 Jan 2003 04:47:29
-1100
Received: from unknown (HELO mxs.perenter.com) (190.44.249.166)
by public.micromail.com.au with NNFMP; Sat, 11 Jan 2003
17:46:47 +0900
Received: from unknown (85.121.248.18) by asx121.turbo-inline.com
with asmt; 12 Jan 2003 02:46:05 +1000
Received: from [98.109.171.85] by external.newsubdomain.com with
local; 12 Jan 2003 12:45:23 -0300
Received: from unknown (HELO qnx.mdrost.com) (205.236.177.234) by
nntp.pinxodet.net with NNFMP; Sun, 12 Jan 2003 09:44:41 -0500
<- SNIP ->
```

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Mail Filtering

- Modern mail software packages have features to filter mail based on:
  - Message headers
  - Message body
  - Sending host, including:
    - IP address
    - DNS lookup
    - SMTP responses
  - Many more ...

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Real-Time Blocking Lists (RBL)

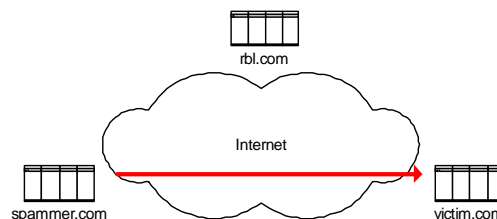
- RBLs provide efficient and consensual blocking of mail hosts known to harbor spammers
- Examples include:
  - <http://www.dnsbl.org>
  - <http://relays.osirusoft.com>
- Caution is advised when choosing your RBL!

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: RBL

1. Spammer starts to send spam to the victim

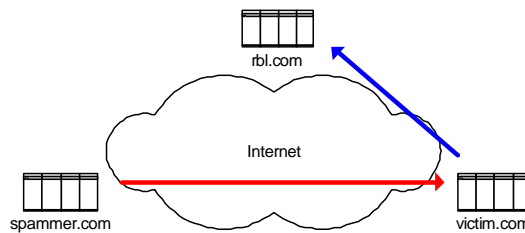


© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: RBL (2)

2. Victim checks with RBL to determine if spammer.com is a known spammer

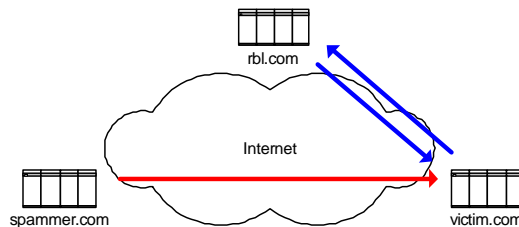


© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: RBL (3)

3. RBL responds that spammer.com is a confirmed spammer

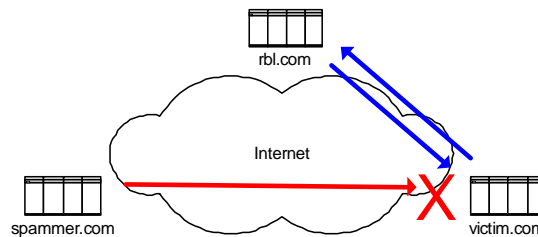


© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: RBL (4)

### 4. Victim blocks mail transmission



© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Regular Expression Matching

- Searches incoming messages for patterns of text that are known to be used by spammers
- Improper sensitivity levels may miss spam or mark legitimate messages as spam
- Very commonly used method

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: Regular Expression Matching

- Set up regular expression filter
- Search for the regular expression:
  - “click below”
- What are the implications?

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Regular Expression Matching: Hit Lists

- Method used to avoid mislabelling legitimate messages
- Every regular expression “hit” is associated with some number of “points”
- When a threshold is met, the mail is marked as spam
- Still not a perfect solution ...

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example: SpamAssassin

```
SPAM: ----- Start SpamAssassin results -----
SPAM: This mail is probably spam. The original message has been altered
SPAM: so you can recognise or block similar unwanted mail in future.
SPAM: See http://spamassassin.org/tag/ for more details.
SPAM:
SPAM: Content analysis details: (15.60 hits, 5 required)
SPAM: MSGID_CHARS_WEIRD (1.6 points) Message-Id has characters often found in spam
SPAM: CONSOLIDATE_DEBT (0.8 points) BODY: Consolidate Debt and Credit
SPAM: REFINANCE (0.6 points) BODY: Refinance Home
SPAM: MORTGAGE_OBFU (0.4 points) BODY: Attempt at obfuscating the word "mortgage"
SPAM: EXCUSE_10 (0.3 points) BODY: "if you do not wish to receive any more"
SPAM: CLICK_BELOW (0.3 points) BODY: Asks you to click below
SPAM: EXCUSE_14 (0.2 points) BODY: Tells you how to stop further spam
SPAM: SPAM_PHRASE_08_13 (1.4 points) BODY: Spam phrases score is 08 to 13 (medium)
SPAM: [score: 11]
SPAM: BIG_FONT (0.3 points) BODY: FONT Size +2 and up or 3 and up
SPAM: LINES_OF_YELLING (0.2 points) BODY: A WHOLE LINE OF YELLING DETECTED
SPAM: WEB_BUGS (0.2 points) BODY: Image tag with an ID code to identify you
SPAM: WEIRD_PORT (1.2 points) URI: Uses non-standard port number for HTTP
SPAM: HTTP_WITH_EMAIL_IN_URL (0.3 points) URI: 'remove' URL contains an email address
SPAM: RAZOR2_CHECK (3.9 points) Listed in Razor2, see http://razor.sf.net/
SPAM: RCVD_IN_OSIRUSOFT_COM (0.4 points) RBL: Received via a relay in relays.osirusoft.com
SPAM: [RBL check: found 27.132.217.209.relays.osirusoft.com., type: 127.0.0.6]
SPAM: RCVD_IN_SBL (3.2 points) RBL: Received via SBLeD relay, see http://www.spamhaus.org/sbl/
SPAM: [RBL check: found 27.132.217.209.sbl.spamhaus.org.]
SPAM: X_OSIRU_SPAMWARE_SITE (0.3 points) RBL: DNSBL: sender is a Spamware site or vendor
SPAM:
SPAM: ----- End of SpamAssassin results -----
```

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## It's Hard for Computers to Filter Spam

- Do you have trouble recognizing spam in your inbox?
- Why is it so difficult for computers to do the same thing that you do so easily?

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Text Searches Vs. Language

- Computers search text for specific strings
- People read text and comprehend language
- How do we program a computer to recognize language in terms that it can understand?

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Bayesian Filtering

- New and exciting method to filter spam
- Currently being implemented in many spam filtering packages
- Filters spam based on a statistical analysis of the contents

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Implementing Bayesian Filtering

- Build two collections of mail:
  - Spam
  - Non-spam
- Collections should be at least 4000 messages for accurate results
- Filter breaks apart messages into a collection of tokens and creates a hash

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Combining the Hashes

- Map each token to the probability that an e-mail containing it is spam

```
(let ((g (* 2 (or (gethash word good) 0)))
      (b (or (gethash word bad) 0)))
  (unless (< (+ g b) 5)
    (max .01
      (min .99 (float (/ (min 1 (/ b nbad))
                        (+ (min 1 (/ g ngood))
                          (min 1 (/ b nbad))))))))))
```

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Compare E-Mail to the Token Hash

- As e-mail is received:
  - Separate the e-mail into tokens
  - Compare it to the hash
  - Based on the outcome, mark it appropriately

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Benefits of Bayesian Filtering

- Considers all evidence in the e-mail, both good and bad
- “Good” words contribute as much to the algorithm as “bad” words

© Copyright 2003 – Chesapeake NetCraftsmen, LLC

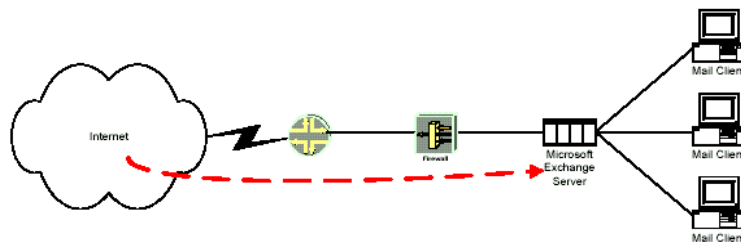


## Spam Filtering Architectures and Examples

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



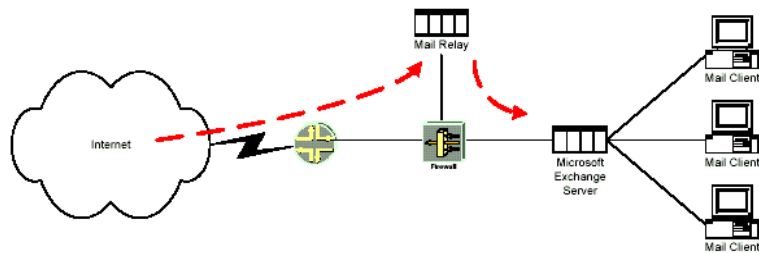
### Example Network: No Filtering



© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example Network: Mail Relay



© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Mail Relays

- Many applications available:
  - Sendmail
  - Postfix
  - Exim
  - Qmail
- All are Open Source applications and comprise the majority of the Internet e-mail backbone

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



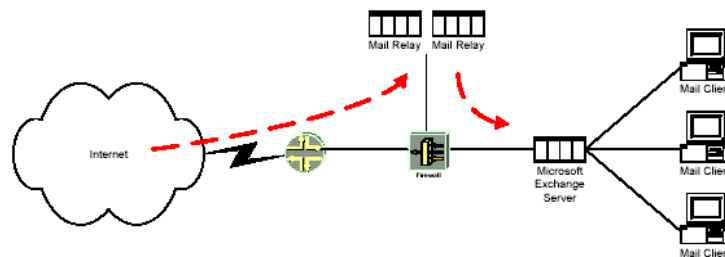
## Preparing the Mail Relay

- The mail relay should:
  - Run on a stable, fault-tolerant operating system
  - Only be running mail applications
  - Be hardened against attack

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example Network: Redundant Relays



© Copyright 2003 – Chesapeake NetCraftsmen, LLC



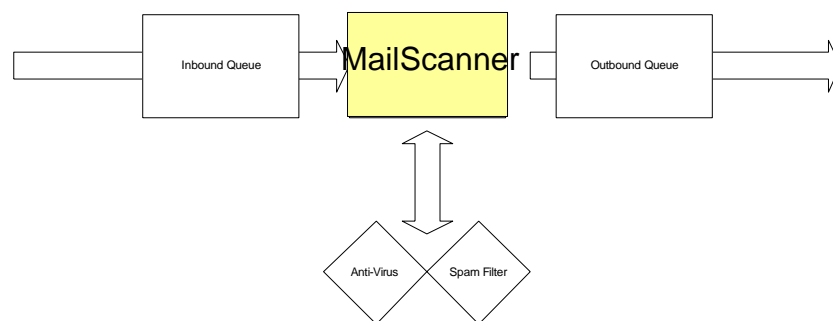
## Relay Filtering Options

- MailScanner
  - <http://www.mailscanner.info>
- Separates incoming and outgoing mail into separate queues
- Runs external anti-virus and spam filtering software to scan incoming mail

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## MailScanner Architecture



© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Anti-Virus Options

- Sophos, McAfee, F-Prot, Command, Kaspersky, Inoculate, Inoculan, Nod32, F-Secure, Panda, RAV, Antivir, ClamAV, Vscan

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



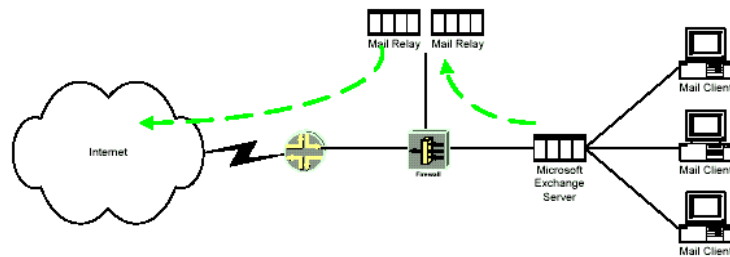
## Spam Filtering Options

- SpamAssassin
  - <http://www.spamassassin.org>
- Supports:
  - RBL
  - Regular expression matching
  - Text analysis
  - Bayesian filtering

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Example Network: Outgoing Mail



© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Summary

- Many methods for fighting unsolicited bulk e-mail are available, and the tools continue to improve.
- If everyone does their part, we can make it very difficult for the spammers!
- It is my hope that this presentation is helpful to you in the future!

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## Your Opportunity to Stump the Presenter

Questions?

© Copyright 2003 – Chesapeake NetCraftsmen, LLC



## In Conclusion

Thank you!

Derek Rogillio

Senior Consultant

Chesapeake NetCraftsmen, LLC

derek (at) netcraftsmen (dot) net

© Copyright 2003 – Chesapeake NetCraftsmen, LLC