

SPAM And SpamAssassin

Alon Altman

Haifa Linux Club



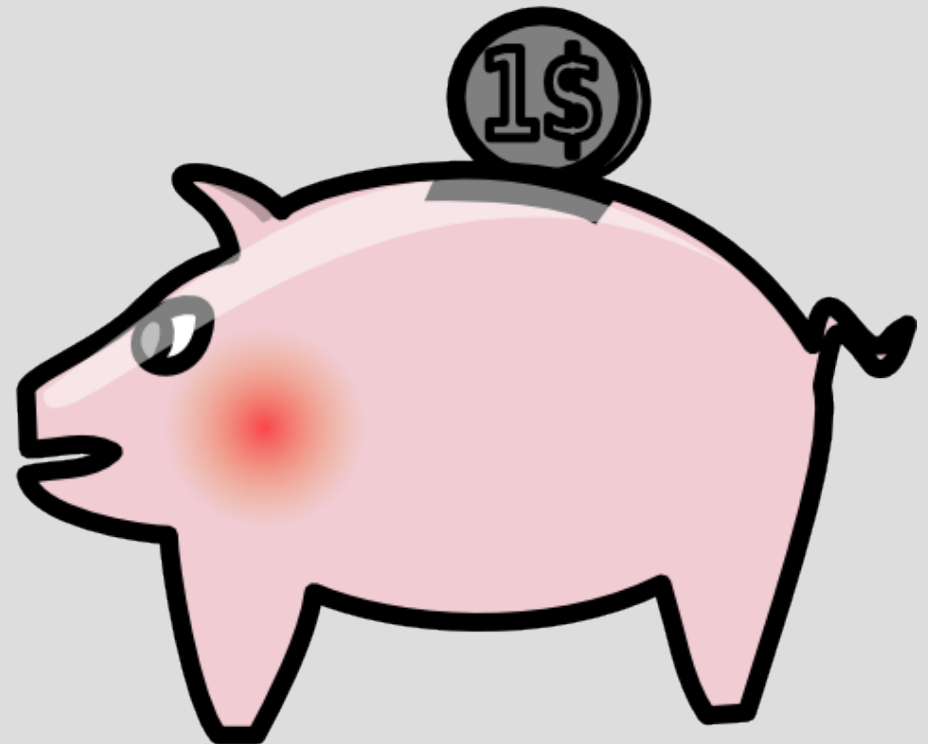
What is SPAM?

- “Unsolicited usually commercial e-mail sent to a large number of addresses.”
 - Merriam-Webster Online Dictionary
- “I know it when I see it.”
 - Justice Potter Stewart (about porn).



Why is this it so hard?

- We all *know* how SPAM looks like, so why is it so hard to filter?



Nigerian 419 Scam

★ Bernard Martin <bernard_martin10@virgilio.it> to list

Attn: Sir/Madam,

{HIGHLY CONFIDENTIAL}
(RE: TRANSFER OF (THIRTY SIX
MILLION SEVEN HUNDRED AND FIFTY NINE
THOUSAND POUNDS STERLINGS)
RE:
TRANSFER OF 36,759,000.00 MILLION POUNDS TO YOUR ACCOUNT.

My name is
Mr. Bernard Martin and I work in the International Operation Department
in a Bank here in London. I feel quite safe dealing with you in this
important business. Though, this medium (Internet) has been greatly
abused, I choose to reach you through it because it still remains the
fastest medium of communication. However, this correspondence is
unofficial and private, and it should be treated as such.

At first I
will like to assure you that this transaction is 100% risk and trouble
free to both parties. WE WANT TO TRANSFER OUT MONEY FROM OUR BANK HERE
IN LONDON. THE FUND FOR TRANSFER IS OF CLEAN ORIGIN. THE OWNER OF THIS
ACCOUNT IS A FOREIGNER, a program leader. Until his death, The Late Prime
Minister, Mr. Rafik Hariri, has a huge investment here in the United
Kingdom and all over the world, as a matter of fact he has the sum of
(THIRTY SIX MILLION SEVEN HUNDRED AND FIFTY NINE THOUSAND POUNDS
STERLINGS) in his account here in London which he deposited as a family
valuables. The family do not know about this deposit. I was on a routine

To buy or not to buy? [Spam](#)

★ [Walter <ralph@woodp.com>](#) to pause

[show details](#) Oct 24

[Reply](#) ▾

[All products for your health](#)

Tortured with health problems? You're one click away from healthy life!

An amazing variety of licensed meds at one big store! Click the link and make your first step to constant relief!

[Men's health](#)

- ◆ [V1agra Soft Tabs](#)
- ◆ [V1agra Professional](#)
- ◆ [C1al1s](#)
- ◆ [C1al1s Soft Tabs](#)
- ◆ [Generic V1agra](#)
- ◆ [Lev1tra](#)

[General health](#)

- ◆ [Human Growth Hormone](#)
- ◆ [Hangover Pills](#)
- ◆ [100% Pure Okinawan Coral Calcium](#)
- ◆ [All-Natural Magnesium Oxide](#)
- ◆ [Soothenol](#)
- ◆ [Quick-detox](#)

[Women's health](#)

- ◆ [Nymphomax](#)
- ◆ [Suregasm](#)
- ◆ [Quick Bust](#)
- ◆ [Pheromone perfume for women](#)

[Weight loss](#)

- ◆ [Meridia](#)
- ◆ [Premium Diet Patch](#)
- ◆ [Liposafe](#)
- ◆ [Lipothin](#)

Creative misspellings

Image with SPAM message

Fwd: MORE [Spam](#)

★ Garry Cannon <Glovernelipegloss@wayne.edu> to bmc3 [show details](#) Dec 12 (2 days ago)

INVESTOR ALERT-IDS.M.OB

WATCH IDS.M TRADE AS MASSIVE PR CAMPAIGN BEGINS.

THIS ONE IS SURE TO BE SEEN BY MILLIONS OF INVESTORS!

GET ON THE TRAIN BEFORE IT LEAVES!

ADD IDS.M TO YOUR RADAR ON WEDNESDAY, DEC 13TH.

Market Info

INDUSTRIAL MINERALS

Symbol: IDS.M.OB

Current Price: .16

5-Day-Target: .90

Rating: STRONG Buy.

Latest Press

07/12/06:

Industrial Minerals, Inc.
Enters Final Stage of Process
Optimization

25/11/06:

INDUSTRIAL MINERALS INC
Financial

Irrelevant text for filters

Martian death-machines He looked toward the barbecue pot, expecting it to look like a barbecue pot in the morning light: a barbecue pot and nothing else.

At the very last moment she pivoted away from him and flung the water-pitcher at the door instead, where it shattered as the soup-bowl had the other day.

***Next Saturday, I was standing in front of the theater at noon, although the box office didnt open until one-fifteen and the movie didnt start until two.**

The spray of dried flowers on the coffee-table had overturned; beneath the table, barely visible, lay a dish of crusted custard pudding and a large book.

And much more...

- Use of HTML codes (e.g. `י`) instead of characters.
- Different versions for text and HTML (assuming filter will work on text version).
- Insert personal details of recipient to bypass filters.

The SPAM arms race

- SPAM filtering is a unique problem in AI.
- Moving target
 - SPAM keeps changing.
 - Deliberately designed to overcome filters.
- Spammers have lots of resources.
 - Thousands of compromised Windows machines send spam continuously.
- Spammers do not care if you're interested.

Do YOU have a solution?

Your post advocates a

technical legislative market-based vigilante

approach to fighting spam. Your idea will not work. Here is why it won't work.

- Spammers can easily use it to harvest email addresses
- Mailing lists and other legitimate email uses would be affected
- No one will be able to find the guy or collect the money
- It is defenseless against brute force attacks
- It will stop spam for two weeks and then we'll be stuck with it
- Users of email will not put up with it
- Microsoft will not put up with it
- The police will not put up with it
- Requires too much cooperation from spammers
- Requires immediate total cooperation from everybody at once
- Many email users cannot afford to lose business or alienate potential employers
- Spammers don't care about invalid addresses in their lists
- Anyone could anonymously destroy anyone else's career or business

Specifically, your plan fails to account for

- () Laws expressly prohibiting it
- () Lack of centrally controlling authority for email
- () Open relays in foreign countries
- () Ease of searching tiny alphanumeric address space of all email addresses
- () Asshats
- () Jurisdictional problems
- () Unpopularity of weird new taxes
- () Public reluctance to accept weird new forms of money
- () Huge existing software investment in SMTP
- () Susceptibility of protocols other than SMTP to attack
- () Willingness of users to install OS patches received by email
- () Armies of worm riddled broadband-connected Windows boxes
- () Eternal arms race involved in all filtering approaches
- () Extreme profitability of spam
- () Joe jobs and/or identity theft
- () Technically illiterate politicians
- () Extreme stupidity on the part of people who do business with spammers
- () Dishonesty on the part of spammers themselves
- () Bandwidth costs that are unaffected by client filtering
- () Outlook

and the following philosophical objections may also apply:

- () Ideas similar to yours are easy to come up with, yet none have ever been shown practical
- () Any scheme based on opt-out is unacceptable
- () SMTP headers should not be the subject of legislation
- () Blacklists suck
- () Whitelists suck
- () We should be able to talk about Viagra without being censored
- () Countermeasures should not involve wire fraud or credit card fraud
- () Countermeasures should not involve sabotage of public networks
- () Countermeasures must work if phased in gradually
- () Sending email should be free
- () Why should we have to trust you and your servers?
- () Incompatibility with open source or open source licenses
- () Feel-good measures do nothing to solve the problem
- () Temporary/one-time email addresses are cumbersome
- () I don't want the government reading my email
- () Killing them that way is not slow and painful enough

Introducing: SpamAssassin

- A server-side tool for classification and scoring of SPAM.
- Classifies one or more e-mail messages as SPAM or HAM (not spam).
- Optionally manipulates spam e-mail messages to mark them.
- Spam can then be moved or deleted using standard filters.



Using spamassassin

- Spamassassin is a filter. It reads a message from STDIN and outputs it to STDOUT.
- To use spamassassin:
 - Pipe all your mail through SA
 - Filter the result using your favorite filter using the X-Spam-Status header.

SA and procmail

- The easiest way to use SpamAssassin is with procmail.
- If you do not use procmail already, first install and configure procmail:
 - apt-get install procmail
 - man procmail
- Then, add recipes to your .procmail file for spam filtering.

.procmail configuration

```
:0fw
```

```
| spamassassin
```

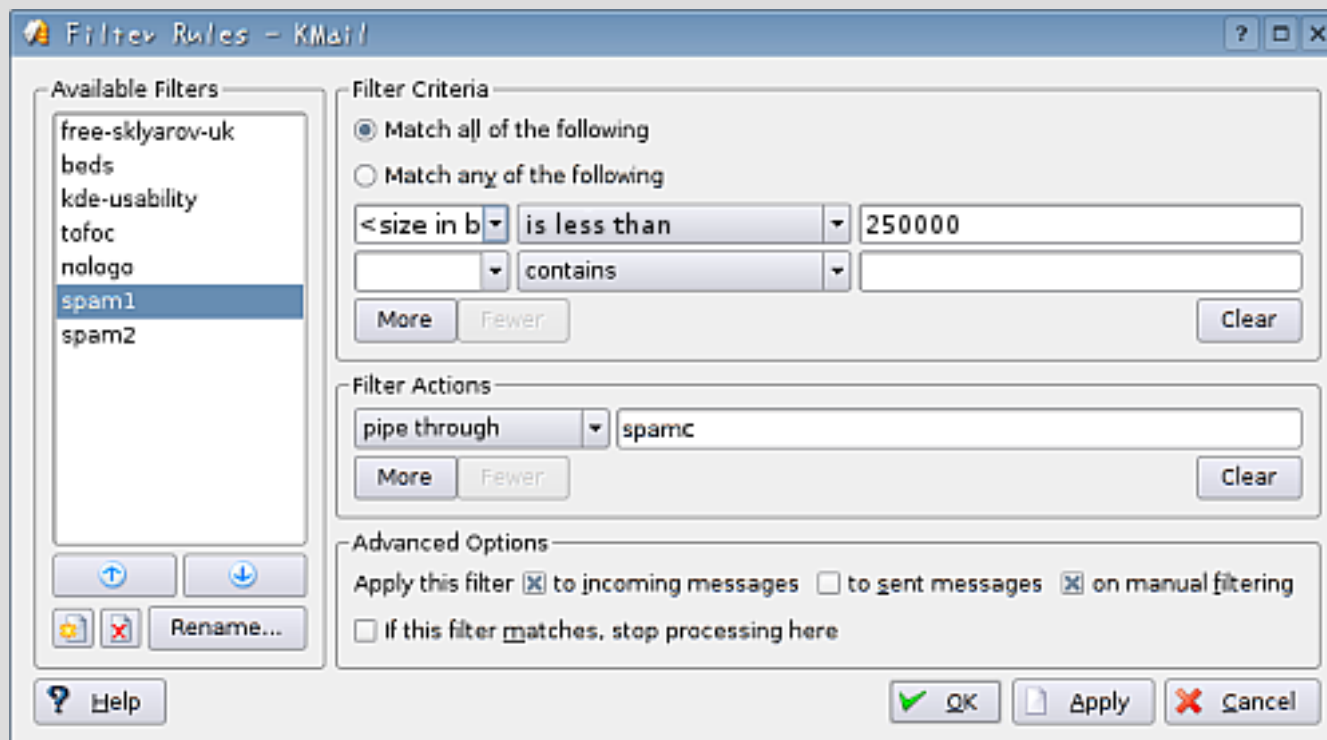
```
:0:
```

```
* ^X-Spam-Status: Yes
```

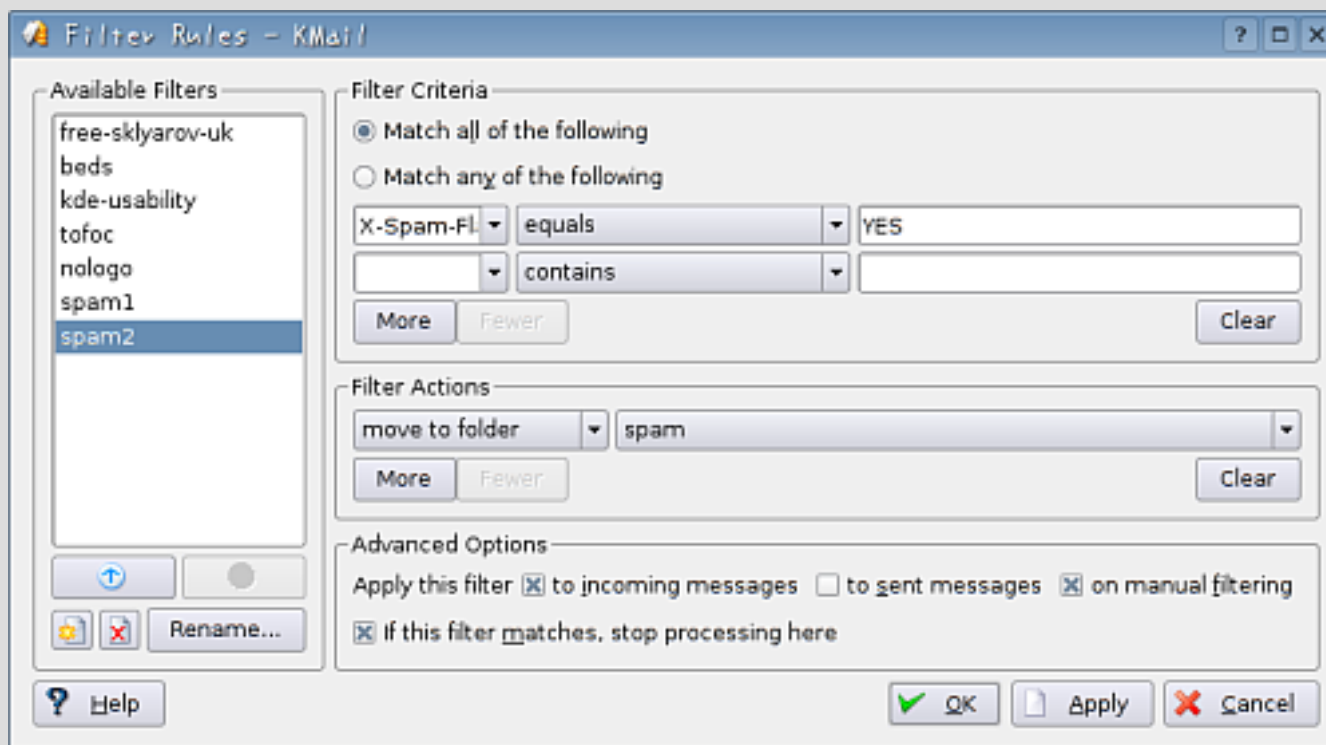
```
junk-mail-folder
```

kmail configuration

- What if you don't own a mail server?
- kmail can be configured to use an external filter as follows:



kmail configuration (2)



Daemon mode

- SA may be heavy to load for each message.
- SA can be used as a daemon (spamd) with a lightweight client (spamc).
- To use SA in daemon mode:
 - Read the spamd README file (yes, I mean it)
 - Enable the SA daemon using your distribution's tools.
 - Use spamc whenever you would have used spamassassin.

How does SA work?

- The basic approach: Try everything!
- SA employs a wide variety of heuristics and services for filtering SPAM.
- All the different methods are combined to a single score.
- Scoring for each rule and the required score are fully customizable.
- We shall now explore some of SA's different methods for filtering SPAM.

Tradeoff

- The SPAM battle is an eternal tradeoff between **false positives** and **false negatives**.
- **False positive** - Legitimate mail classified as SPAM.
- **False negative** - SPAM classified as legitimate mail.
- Use `required_hits` setting to specify your SPAM tolerance. 5 is a good value.

SA Configuration

- SA configuration files define rules and scoring for SPAM filtering as well as general options.
- Standard rule configuration files are in `/usr/share/spamassassin/*`
- System-wide configuration can be placed in `/etc/spamassassin/local.cf`
- User-specific configuration is in `~/.spamassassin/user_prefs`

Testing SA: GTUBE

- GTUBE is the Generic Test for Unsolicited Bulk Email.
- It specifies a test string that every spam filter should classify as spam.
- To test your SA configuration, send yourself a message with the following test (with no spaces or line breaks):

```
XJS*C4JDBQADN1.NSBN3*2IDNEN*GTUBE-  
STANDARD-ANTI-UBE-TEST-EMAIL*C.34X
```

SpamAssassin Filters

Message text filters

Content Filtering

- **Basic idea:** Match phrases commonly used in SPAM.
- **Examples:**
 - HTML content tests
 - URL tests
 - `body __FRAUD_EZY /\b(?:of|the) late president\b/i`
 - `body DRUG_ED_CAPS /\bCIALIS|VIAGRA/`
 - `body FREE_PORN /\bfree(?:porn|xxx)/i`
 - `body 100_PERCENT /100% GUARANTEED/i`
 - `body EXCUSE_23 /you have provided permission/i`

Hebrew & Israeli SPAM

- SA's default filters do not contain Hebrew phrases or Israeli websites.
- Special rules for Hebrew and Israeli SPAM have been written by Ilan Asic and Gal Ben-Haim.
- Download from:
 - <http://www.deltaforce.net/hebrewspam/>

Metadata filtering

- **The idea:** Use spammers' attempts to evade filters against them.
- Example rules:
 - Multipart message mostly text/html MIME
 - Message text disguised using base64
 - Bulk email software fingerprint
 - From: starts with many numbers
 - Message-Id is fake
 - Date in (distant) past/future
 - Message body is 75-100% uppercase
 - Message includes Microsoft executable

Language-based rules

- **The Idea:** Messages in foreign languages must be spam.
- Example rules:
 - Character set or message content indicates a foreign language
 - Headers have too many raw illegal characters
- **Important:** Specify your language(s) in the configuration file:
 - `ok_locales en he`

Bayesian Filtering

- Made popular by Paul Graham's article "A plan for Spam" (2002).
- Was subject of academic research much earlier
 - In fact, I've implemented a Bayesian SPAM filter before that paper was written.
 - *ifile* implemented general Bayesian filtering way back in 1996.

What's Bayesian Filtering?

- Bayes' theorem gives a method for reversing probabilities:

$$Pr(A|B) = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}$$



- Applied to SPAM filtering, we get:

$$Pr(\textit{Spam}|\textit{Message}) \propto Pr(\textit{Message}|\textit{Spam})$$

Idea of Bayesian Filtering

- **Train** a classifier using examples of both SPAM and HAM.
 - Training identifies words that appear a lot in either HAM or SPAM.
- **Apply** the classifier to get a probability or a score for unknown messages.
 - Reduce SPAM score for messages likely to be HAM.
 - Increase SPAM score for messages likely to be SPAM.

Bayesian Training

- In order to teach SA, simply pipe individual messages or whole folders to sa-learn with flag --ham or --spam.
- SA also automatically learns by believing its own classifications.
 - If a message has been wrongly classified, you can tell spamassassin to report it (-r) as spam or revoke it (-k), or pipe it to sa-learn as above.

Message text Filtering: Why not?

Your post advocates a

(*) technical () legislative () market-based () vigilante approach to fighting spam. Your idea will not work.

Here is why it won't work.

- (*) Mailing lists and other legitimate email uses would be affected
- (*) Users of email will not put up with it
- (*) Many email users cannot afford to lose business or alienate potential employers


Specifically, your plan fails to account for

- (*) Eternal arms race involved in all filtering approaches
- (*) Extreme profitability of spam
- (*) Dishonesty on the part of spammers themselves
- (*) Bandwidth costs that are unaffected by client filtering
- (*) Outlook

and the following philosophical objections may also apply:

- (*) We should be able to talk about Viagra without being censored

Digest Based Approaches

- **The Idea:** SPAM gets sent to many people at once. Several users report the SPAM, all others filter.
- Use hash of message instead of actual text.
- Implementations supported by SA:
 - Vipul's Razor2 & Pyzor 
 - Distributed Checksum Clearinghouse(DCC)
- The -r (report) and -k (revoke) switches of SA also report/revoke the message in the digest servers.

Digest-based Filtering: Why not?

Your post advocates a

(* technical () legislative () market-based () vigilante approach to fighting spam. Your idea will not work.

Here is why it won't work.

- (* Mailing lists and other legitimate email uses would be affected
- (* It will stop spam for two weeks and then we'll be stuck with it
- (* Users of email will not put up with it
- (* Microsoft will not put up with it
- (* Anyone could anonymously destroy anyone else's career or business

Specifically, your plan fails to account for

- (* Eternal arms race involved in all filtering approaches
- (* Extreme profitability of spam
- (* Dishonesty on the part of spammers themselves
- (* Bandwidth costs that are unaffected by client filtering

and the following philosophical objections may also apply:

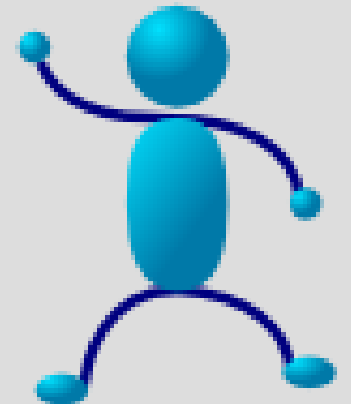
- (* Blacklists suck
- (* Why should we have to trust you and your servers?

SpamAssassin Filters

Network-based Filters

Message Relay tests

- **The idea:** Try to determine if message was sent using nonstandard path, or headers were forged.
- Example Rules:
 - Passed through trusted hosts only via SMTP
 - Host HELO did not match reverse DNS
 - HELO from dynamic IP
 - Message routed round-the-world



DNS Blocklists

- **The Idea:** Check IP addresses in mail headers against lists of known spammers, dailup accounts, and SPAM supporters.
- Supported Blocklists:
 - **IP Addresses:** NJABL, SORBS, Spamhaus, rfc-ignorant, CompleteWhois, DSBL, RHSBL, SpamCop, MAPS
 - **URLs:** Spamhaus, SURBL

S  **RBS**



DNSBLs - Why not?

Your post advocates a

(* technical () legislative () market-based () vigilante approach to fighting spam. Your idea will not work.

Here is why it won't work.

- (* It is defenseless against brute force attacks
- (* Users of email will not put up with it
- (* Many email users cannot afford to lose business or alienate potential employers
- (* Anyone could anonymously destroy anyone else's career or business

Specifically, your plan fails to account for

- (* Armies of worm riddled broadband-connected Windows boxes
- (* Joe jobs and/or identity theft
- (* Bandwidth costs that are unaffected by client filtering

and the following philosophical objections may also apply:

- (* Blacklists suck
- (* Why should we have to trust you and your servers?
- (* Incompatibility with open source or open source licenses

DomainKeys and SPF

- **The Idea:** Authenticate the sender address of the message.
- **Purpose:** Avoid Joe Jobs - Spammers sending mail using other people's addresses as "From" addresses.
- **Solution:**
 - **SPF** - Use DNS to mark allowed senders for messages
 - **Domain Keys** - Mail relay cryptographically signs outgoing mail.

Joe Jobs

Date: Fri, 15 Dec 2006 08:57:04 +0800
From: Essie T. Travis <dlg@8ln.org>
To: bakary.ba@kalix.fr
Subject: This is a magazine about romance.



All replies and bounces are directed to my address!

Use Authenticated e-mail

- Authenticated e-mail does not mean it isn't SPAM!
- It only means you can trust the sender address as real.
- Failed authentication is a sign of SPAM.
- Authenticated addresses could be checked against a whitelist.

Sender Authentication: Why not?

Your post advocates a

(*) technical () legislative () market-based () vigilante approach to fighting spam. Your idea will not work.

Here is why it won't work.

- (*) Mailing lists and other legitimate email uses would be affected
- (*) It will stop spam for two weeks and then we'll be stuck with it
- (*) Users of email will not put up with it
- (*) Requires too much cooperation from spammers

Specifically, your plan fails to account for

- (*) Armies of worm riddled broadband-connected Windows boxes
- (*) Extreme profitability of spam
- (*) Joe jobs and/or identity theft
- (*) Dishonesty on the part of spammers themselves

and the following philosophical objections may also apply:

- (*) Whitelists suck
- (*) Why should we have to trust you and your servers?
- (*) Feel-good measures do nothing to solve the problem

SpamAssassin Filters

Methods to mark HAM

HashCash

- **The Idea:** Perform a hard computational task to ensure delivery of your mail.
- Using HashCash, messages can be marked with a special “stamp” that will contain proof of spending CPU time for a specific message and recipient.
- Spammers will require too much CPU time to send all their e-mails.
- SA assigns negative values to messages with proper HashCash headers.

Sender Authentication: Why not?

Your post advocates a

(*) technical () legislative () market-based () vigilante approach to fighting spam. Your idea will not work.

Here is why it won't work.

- (*) Mailing lists and other legitimate email uses would be affected
- (*) It is defenseless against brute force attacks
- (*) Users of email will not put up with it

Specifically, your plan fails to account for

- (*) Willingness of users to install OS patches received by email
- (*) Armies of worm riddled broadband-connected Windows boxes
- (*) Bandwidth costs that are unaffected by client filtering
- (*) Outlook

and the following philosophical objections may also apply:

- (*) Whitelists suck
- (*) Sending email should be free

HABEAS

- **Basic Idea:** Copyrighted and trademarked Haiku allowed only in non-SPAM mail.
- SA gives a negative SPAM score to messages which contain the following headers:

X-Habeas-SWE-1: winter into spring

X-Habeas-SWE-2: brightly anticipated

X-Habeas-SWE-3: like Habeas SWE (tm)

X-Habeas-SWE-4: Copyright 2002 Habeas (tm)

X-Habeas-SWE-5: Sender Warranted Email (SWE) (tm). The sender of this

X-Habeas-SWE-6: email in exchange for a license for this Habeas

X-Habeas-SWE-7: warrant mark warrants that this is a Habeas Compliant

X-Habeas-SWE-8: Message (HCM) and not spam. Please report use of this

X-Habeas-SWE-9: mark in spam to <<http://www.habeas.com/report/>>.

HABEAS - Why not?

Your post advocates a

() technical (*) legislative () market-based () vigilante approach to fighting spam. Your idea will not work.

Here is why it won't work.

- (*) No one will be able to find the guy or collect the money
- (*) Users of email will not put up with it
- (*) Requires too much cooperation from spammers

Specifically, your plan fails to account for

- (*) Open relays in foreign countries
- (*) Jurisdictional problems
- (*) Armies of worm riddled broadband-connected Windows boxes
- (*) Extreme profitability of spam
- (*) Joe jobs and/or identity theft
- (*) Dishonesty on the part of spammers themselves

and the following philosophical objections may also apply:

- (*) SMTP headers should not be the subject of legislation
- (*) Sending email should be free
- (*) Feel-good measures do nothing to solve the problem

For Further Reading

- <http://www.spamassassin.org/>
- “A plan for SPAM” - Paul Graham
- <http://www.pyzor.sf.net/>
- <http://www.opendpf.org/>
- <http://www.hashcash.org/>
- RFCs 2045-2049, 2821