

# Spam Conference

January 17, 2003  
Massachusetts Institute of Technology  
Cambridge, MA

<http://spamconference.org/>  
<http://paulgraham.com/spam.html>

## Slide 1

### Notes for slide 1:

- On January 17th I attended a “Spam Conference” at MIT.
- So, you may ask, what is spam and why a conference about it.
- Or, you may already know and be spending a portion of your life handling spam.

## Spam Conference 2003

- Bulk, Unsolicited Commercial Email (Not the food)
- Paul Graham, *A Plan for Spam*  
(<http://paulgraham.com/spam.html>)
- Bayesian Filtering
- Intended: small gathering, anti-spam methods
- Result: 560 participants

### Slide 2

#### Notes for slide 2:

- This is the Bulk, Unsolicited Commercial Email kind, not the processed food substance.
- This conference was organized by Paul Graham of MIT after he wrote a short article *A Plan for Spam*, in which he proposed Bayesian Filtering to sort out commercial, or otherwise unwanted, email messages. (Turns out to be an old idea, but more on that later.)
- This was intended to be a small gathering to compare anti-spam methods.
- It grew to 560 participants, was covered by *Wired*, *Computer World*, and other news sources.
- Alas, this meant it was very hit-and-run: show up, people talk, little opportunity for networking.
- Still, talks were informative, G&A insightful, and the explosive growth of conference interest a sign of the explosive growth, and impact, of Spam.

## What is Spam?

- Unsolicited
- Commercial
- Bulk
- Email
- Usenet, Pop-up alerts, . . . .

### Slide 3

#### Notes for slide 3:

There are many shades of spam, as the term is commonly used. But, at this conference, speakers who offered a definition emphasized:

1. Unsolicited—nobody asked for this (opt-out, versus opt-in).
2. Commercial—but, not all “spam” is commercial (distinction between commercial—selling something, and solicitation—selling something other than goods and services).
3. Bulk—not a single or small group of recipients, instead 10's of thousands or millions.
4. Most attention on email, but there are other forms of spam:
  - (a) Original spam was to newsgroups—still a problem,
  - (b) Pop-up spam via windows message service.
  - (c) More vectors will be exploited if email is denied.

### **Overview of Talks:**

- Spam methods
- Filtering methods
- Meta-filtering
- Impact of Spam

### **Slide 4**

#### **Notes for slide :**

- Spammer methods (John Graham-Cumming, POPFile)
- Filtering methods (majority of talks)
- Meta-filtering (what works, what doesn't work)
- Impact of Spam on ISP, Users, The Internet.

### **Spam Methods:**

- Images and URLs
- Splitting keywords
- HTML messages
- Open relays, proxy relays
- FreeISP accounts
- Retaliatory spamming

#### **Slide 5**

#### **Notes for slide :**

- Images and URLs
- Splitting keywords
- HTML messages
- Open relays, proxy relays
- Free, or low-cost, ISP accounts
- Retaliatory spamming (threatened lawsuits)

### **Filtering Methods:**

- Blacklists  
(<http://relays.osirusoft.com/>)
- Hueristic  
(<http://www.spamassassin.org/>)
- Statistical  
(<http://bogofilter.sourceforge.net/>)
- Intelligence  
(<http://www.messagelabs.com/>)

#### **Slide 6**

#### **Notes for slide :**

- Blacklists (IP or host based), numerous available  
(<http://relays.osirusoft.com/cgi-bin/rbcheck.cgi>)
- Hueristic  
(<http://spamassassin.taint.org/>)
- Statistical  
(<http://bogofilter.sourceforge.net/>)
- Intelligence (human or AI)  
(<http://spamfence.net/>, <http://www.messagelabs.com/>)

### **Meta-Filtering:**

- Statistical Language Processing
- Natural Language Processing
- Numerical Methods
- Comparison of methods
- Real Data?
- Hard AI (Active Adversary)

#### **Slide 7**

#### **Notes for slide :**

- Statistical Language Processing (No learning set is big enough)
- Natural Language Processing (Slow and difficult)
- Numerical Methods (Linear—LMMSE, as good or better than GA, NN, etc.)
- Comparison of methods
  - Heuristic versus Statistical counter claims
  - Personalized is always better.
  - Few studies, often incomplete data.
  - Fast linear better than slow GA.
- What do we know, when are we guessing
- This is a hard AI problem (AI with active adversary)

### **Spam, Spammers and the Internet:**

- No Economic Incentive
- Loss of Trust
- Skimming Internet Profits
- Organized Crime?
- Denial of Service.
- Mailing Lists

#### **Slide 8**

#### **Notes for slide :**

- Barry Shein (CEO, TheWorld)
  - no economic incentive against spam.
  - economic incentive for spam
  - loss of trust (did you sell my email address?)
  - Skimming Internet Profits
  - Spam is organized crime (Jon Praed, Esq. concurs).
  - Indistinguishable from DOS.
- Legal challenges to spam (already illegal, like bank robbery)
- David Berlind, C-NET mailing lists (economic incentives will change the Internet)

### **Legal Action Against Spammers:**

- Emphasized Bulk-UCE
- Already Illegal  
(<http://www.spamlaws.com/>)
- Can Always Trace
- Injunctions, Damages, Bankruptcy
- Spammers have no Money

#### **Slide 9**

#### **Notes for slide 9:**

Jon Praed:

- Emphasized Bulk-UCE: Challenged spammers in court to show recipient has requested spam—nothing ever produced.
- Spam is already illegal:
  - Trespass to Chattels,
  - State Statutes,
  - Acceptable Use Policy,
  - Computer Fraud and Abuse Act.
- Can always trace spammers.
- Spammers operate like an illegal organization.
- Spamming is the first successful job.
- Injunctions, Damages, Bankruptcy
- Never have money.
- Getting more organized.

**Implications:**

- Spam will get worse.
- If it ever gets better.
- Spam is Changing the Internet.
- Unpleasant Topic, WOMBAT.
- Spammers Do Not Care.
- “We have met the enemy and they is us.”, Walt Kelly.

**Slide 10****Notes for slide :**

- It will get worse, before it gets better.
- If it ever gets better.
- The Internet is changing because of spam (David Berlind: Blacklists, filters, trust).
- It will change more (spam until proven ham).
- Unpleasant topic, waste of money, brains and time,
- But, spammers do not care.
- “We have met the enemy and they is us.”, Pogo, Walt Kelly. Everybody has something to tell us. Corporations, Non-profits, Protesters, Faster-Than-Light and Anti-Gravity. Opt-out is a loose.