

# Using Uncensored Communication Channels to Divert Spam Traffic

Benjamin Chiao and Jeffrey MacKie-Mason

Support from the NET Institute, NSF grants IIS-0414710 &  
IGERT-0114368. (Apr 2008)

# Today's Talk

→ Quick Overview

Deeper Background

Theory (Big Picture)

Theory (Details)

# Why Do We Care?

US businesses spent \$10 billion a year on spam (Source: Washington Post). If the “Spam Problem” is solved:

- ▶ You spend less time sorting spam out
- ▶ Advertisers revive email advertising
- ▶ Email service providers spend less money blocking spam
- ▶ Customers spend less money, enjoy more informative ads

# What is an Uncensored (Open) Channel?

- ▶ Keep inbox filters
- ▶ No filters in a special folder
- ▶ Guarantee delivery of tagged messages to this folder

# Hypothetical Design

Sender's side


Send Save Now Dis

**From:** Benjamin Hak-Fung Chiao <ben...>

**To:** "Jeffrey MacKie Mason" <jmm@...>

[Add Cc](#) | [Add Bcc](#)

**Subject:** [uncensored] Rolex, Viagra, etc.

 [Attach a file](#)

Recipient's side



[Inbox \(12\)](#)

[Uncensored \(100\)](#)

[Sent Mail](#)

[Drafts \(5\)](#)

[Spam](#)

# Infrastructure

- ▶ Almost none
- ▶ Fully reversible if it doesn't work

# Why does it Work? Basic Idea is...

## Existing Mechanisms

- ▶ ↑ sending cost. e.g. filters, legal punishment, email stamps

## Our Proposal

- ▶ ↓ benefits of flooding the inbox
  - ▶ by moving out customers
  - ▶ so senders send less to inbox

# Main Results

After implementing the open channel:

- ▶ Recipients enjoy more informative ads and lower prices
- ▶ Senders incur lower volume generation and disguise cost
- ▶ Advertisers incur less advertising costs
- ▶ Spam is only sent to the open channel

# Is this Mechanism Unique?

- ▶ Flood management
- ▶ “Coffee shops” in Amsterdam
- ▶ TV shopping channels

# Secondary Issues

- ▶ Induce more socially objectionable goods?
- ▶ No filtering too extreme?
- ▶ Will open channel be flooded with advertisements?

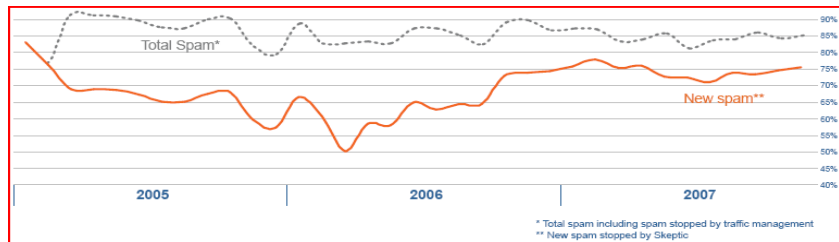
Quick Overview  
→ **Deeper Background**  
Theory (Big Picture)  
Theory (Details)

# What is Spam?

- ▶ No consensus definition in the literature
- ▶ We are not defining it here
- ▶ We target one sub-species usually blocked by filters:
  - ▶ “Unsolicited Commercial Mass” (UCM) mail
  - ▶ Rule out some phishing, malicious mail
- ▶ For convenience today, we use spam for UCM

# How Serious is the Problem in Terms of Volume?

Fig 1: Spam volume over total volume of email on the network



Source: Sophos (2007)

## Types of Spam [Evelt (2006)]:

- ▶ Medication (40%), adult content (19%), others (41%)

# What Makes UCM Advertising Special?

- ▶ Informative v. Persuasive Advertising (Taylor 1934)
  - ▶ UCM reverses common wisdom: Higher portion of informative advertising
  - ▶ We rule out some persuasive advertising

# Sample Advertising Email

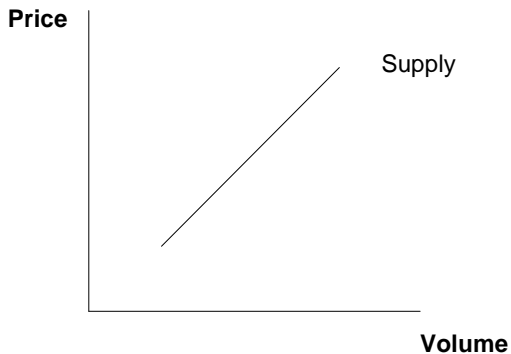
*from: Toby Hager <benjaminchapman@gmail.com>  
to: benjaminchapman@gmail.com,  
date: Jan 15, 2008 5:29 AM  
subject: Repl1ca watch! vpiE  
repl1ca of w4tches  
[http://www.bbeuyye.com/  
vv0nfgoX2MTFC](http://www.bbeuyye.com/vv0nfgoX2MTFC)*

# What Makes UCM Advertising Special?

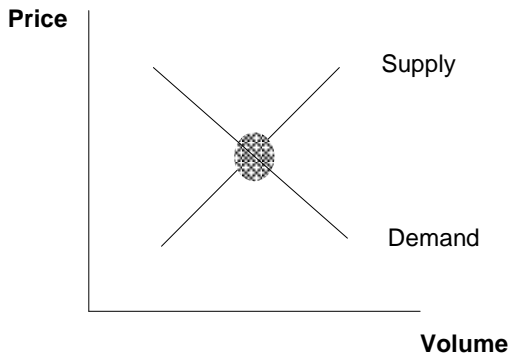
## ▶ Push v. Pull Advertising

- ▶ MessageLabs (2005): 30% of domains die in 24 hrs
- ▶ UCM sites move rapidly, need to push to customers information of new locations
- ▶ Advertisers hire senders (e.g. [send-safe.com](http://send-safe.com)) to get it done

# Is there a Real Demand? Revealed Preference Argument



# Is there a Real Demand? Revealed Preference Argument

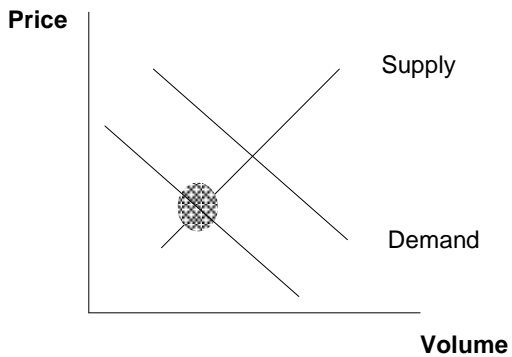




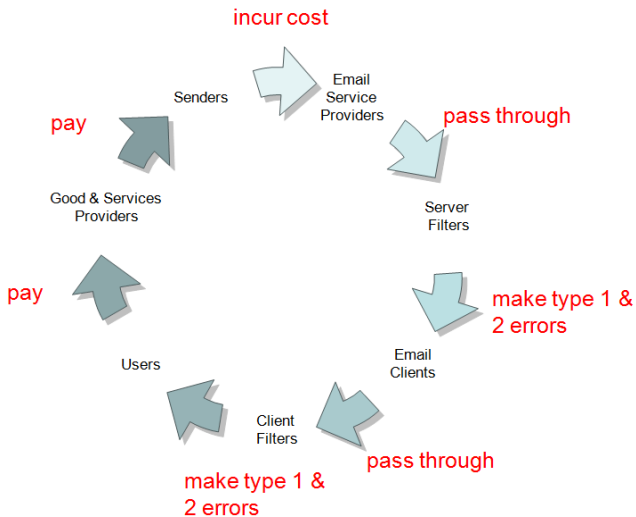
# But how about UCM?

## Real v. Spurious Demand for UCM

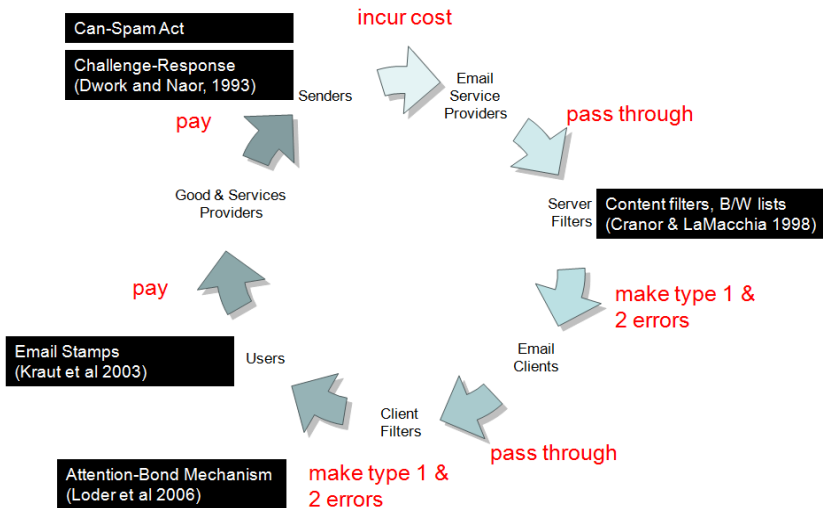
- ▶ Rule out some spurious: malicious (e.g. virus), deceptive (e.g. eBay phishing) email



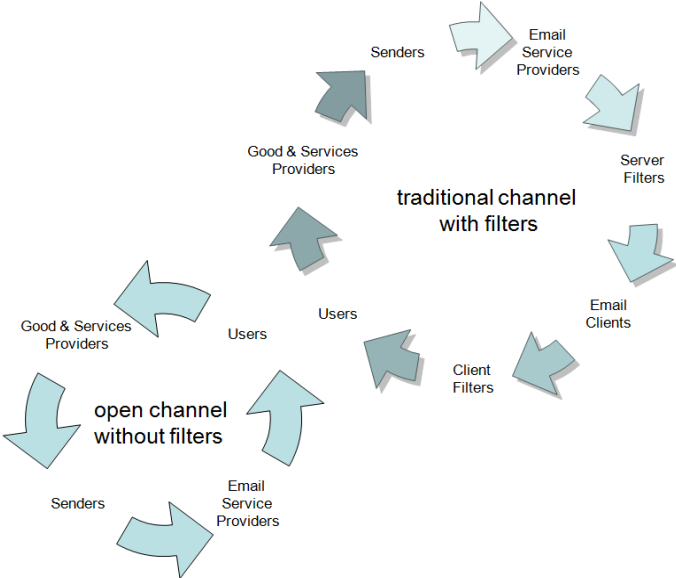
# An Ecosystem of Email Advertising



# Prior Mechanisms



# Our Mechanism



## Broader Implications from this Instance

- ▶ Ours is a way to recreate a free market (e.g. Google AdWords, stamps, bonds)
- ▶ They may be other ways to reduce benefits of sending to inbox

# Will it Work in Practice?

- ▶ The ultimate test is an empirical question
- ▶ We specify some key conditions that it will work in theory

Quick Overview

Deeper Background

→ Theory (Big Picture)

Theory (Details)

# Two Channels

- ▶ (o)pen channel
- ▶ (c)ensored channel

# Recipients

- ▶ Recipient chooses whether to opt-in for the open channel to max utility
- ▶ Utility depends on:
  - ▶ (+) Goods consumed
  - ▶ (-) UNwanted mail received (Type I errors)
  - ▶ (-) Wanted mail NOT received (Type II errors)
- ▶ High-types want some UCM. Low-types don't.

# Senders

Revenue:

- ▶ Advertising charge  $\times$  mail received

Cost:

- ▶ Constant marginal volume generation cost
  - ▶ 60% of UCM are sent by zombies, which are killed constantly
- ▶ Negative transparency cost

# Advertisers

## Revenue:

- ▶ Sales price  $\times$  response rate  $\times$  mail received
- ▶ Response rate depends on:
  - ▶ the ratio of demanders
  - ▶ the transparency level

## Cost:

- ▶ Constant marginal production cost
- ▶ Constant marginal advertising cost paid to senders

## Main Results Revisited

Besides the status-quo, the only pure-strategy Nash is: (i) only high types use the open channel and UCM is sent only to the open channel, (ii) everyone is better off.

## Some Intuition: Why not deviates?

- ▶ Question: If no sender sends to the censored channel, will I make more profit if I send some there?
- ▶ Answer: You are advertising for your competitors. Other senders can always undercut your prices in the open channel because they don't need to incur disguise and costly volume generation cost. If the search cost in the open channel is low enough, customers will buy from there instead.

Quick Overview  
Deeper Background  
Theory (Big Picture)  
→ Theory (Details)

# Fooling A Filter

## ▶ Notations

- ▶  $N$  is mail before filtering
- ▶  $f \in [1, \infty)$  is the filter strength
- ▶  $t \in [\frac{1}{f}, 1]$  is the transparency level
- ▶  $n \equiv \frac{N}{tf}$  is mail received by recipients

## ▶ Examples

- ▶  $N = 10, f = 2, t = 1$  (fully transparent):  $n = 5$
- ▶  $N = 10, f = 2, t = \frac{1}{f}$  (fully disguised):  $tf = 1, n = 10$

# Preference for UCM

- ▶ high type recipients
  - ▶ like  $1 - \epsilon$  portion
  - ▶ have positive money to spend on UCM goods
- ▶ low type recipients
  - ▶ dislike all of them
  - ▶ have no money

# Recipients

Assume that the closed channel is essential, recipient  $r$  chooses whether to opt-in for the open channel to maximize:

$$U_r(v_r^g, v_r^I, v_r^{II})$$

Recipients	Vol. of Gds	Type 1 errors	Type 2 errors
Low	zero	UCM received	zero
High	$\frac{\text{money}}{\text{cheapest price across channels}}$	$\epsilon \times \text{UCM received}$	$(1 - \epsilon) \times \text{UCM not received}$

# Senders' Problem

Each sender  $s$  chooses  $(N_s^o, N_s^c, t_s^c)$  to maximize:

$$\pi_s(N_s^o, N_s^c, t_s^c) = \text{Revenue} - \text{Cost} \quad (1)$$

s.t.

$$t_s^c \in \left[\frac{1}{f^c}, 1\right], N_s^o, N_s^c \geq 0 \quad (2)$$

# Advertisers' Problem

Each advertiser  $a$  chooses  $(n_a^o, n_a^c)$  to maximize:

$$\pi_a(n_a^o, n_a^c) = \text{Revenue} - \text{Cost} \quad (3)$$

s.t.

$$n_a^o, n_a^c \geq 0 \quad (4)$$

# Representative Agents Modeling

- ▶ Three agents: Recipient, Sender, and Advertiser
- ▶ Limitations
  - ▶ We have not formally modeled Bertrand or Cournot type competitions but some implications are subsumed in the zero profit conditions

## When the Open Channel is Available for Opt-in...

## Sender and Advertiser's Best Responses

Note: Low types actions do not affect sellers best responses in terms of signs (not magnitudes) of mail volumes.

Mail Volume in Channel:		High Types Recipients Opt-in for the Open Channel	
Open	Censored	Yes	No
0	0	$\pi_{undefined}$	$\pi_{undefined}$
0	+	$\pi_{waste-customers}$	$\pi_{status-quo}$
+	0	$\pi_{efficient}$	$\pi_{waste-mail}^1$
+	+	$\pi_{duplicate-effort}$	$\pi_{waste-mail}^2$

# Recipient's Best Responses

Case  $(+,0) \implies$  The best response for low types is no and for high types is yes.

Proof.

Recipients	Volume of Goods ( $v_r^g$ )	$v_r^I$	$v_r^{II}$
Low types	zero anyway	$\uparrow$	zero anyway
High types	$\uparrow$ (MU $\uparrow$ a lot as $v_r^g \rightarrow 0$ )	$\uparrow$	unchanged

Table: Utility Change Using the Open Channel



Case  $(0,+)$   $\implies$  The best responses for each recipient are yes and no:

Proof.

Trivial. □

## Main Results Revisited (Nash Equilibria)

Proposition 1: Besides the status-quo, the only symmetric pure-strategy Nash is: Only high types use the open channel and UCM is sent only to the open channel.

Proof.

By intersection of best responses across agents. □

Now, Compare with the Status-Quo of No Open Channel

## Main Results Revisited (Welfare)

### Key assumptions used in:

- ▶ Proposition 3A
  - ▶ assumes free disposal of UCM in the open channel for high-types ( $\epsilon = 0$ ):
    - ▶ e.g. search engines within the open channel
- ▶ Proposition 3B
  - ▶ uses a weaker condition than free disposal:
    - ▶ Proposition 2: less UCM received after implementation

## Proposition 3A (Welfare)

If Prop 1 holds and free disposal exists ( $\epsilon = 0$ ), everyone is weakly better off.

**Proof.**

(i) Advertisers and senders make zero profit anyway.

(ii) Recipients:

H-type	$v_r^g$	$v_r^l$	$v_r^{ll}$
After	$\uparrow$	unchanged	0
Reason	lower price	0 anyway (by $\epsilon = 0$ )	no filter

L-type	$v_r^g$	$v_r^l$	$v_r^{ll}$
After	0	$\downarrow$	0
Reason	0 anyway	UCM only in open ch.; not opt in (by P1)	0 anyway



## Proposition 3B (Welfare)

If Prop 1 holds and UCM  $\downarrow$  (by Prop 2), everyone is weakly better off.

**Proof.**

(i) Advertisers and senders make zero profit anyway.

(ii) Recipients:

H-type	$v_r^g$	$v_r^l$	$v_r^{ll}$
After	$\uparrow$	$\downarrow$	0
Reason	lower price	UCM vol. $\downarrow$ (by P2)	no filter

L-type	$v_r^g$	$v_r^l$	$v_r^{ll}$
After	0	$\downarrow$	0
Reason	0 anyway	UCM only in open ch.; not opt in (by P1)	0 anyway

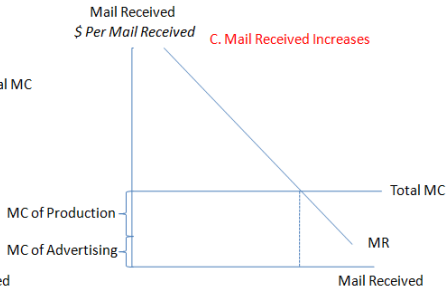
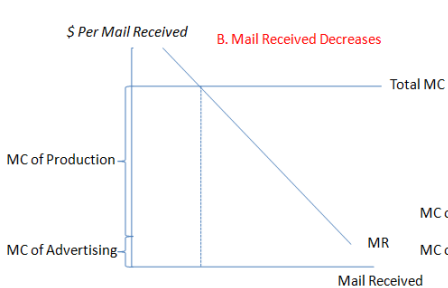
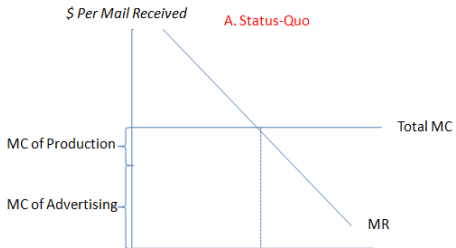


[OPTIONAL] Show Proposition 2: Less UCM after implementation

## [OPTIONAL] Market Clearing Condition for Mail Volume

$$\frac{\text{Expenditure}}{\text{price}} = \text{Goods Volume} = \text{Email Volume} \times \text{Response Rate}$$

- ▶ Our results says “Goods Volume” and “Response Rate”  $\uparrow$
- ▶ But the market condition says “Email Volume” could  $\uparrow / \downarrow$



## Mechanism Review

- ▶ The open channel enables the recipients to enjoy more informative ads and lower sale prices
- ▶ The sale prices are lowered (∵ sending and advertising costs are lower)
- ▶ Sending cost is lowered because there is no need to:
  - ▶ hire programmers to disguise messages
  - ▶ hijack computers to send messages
  - ▶ send more to get a message received (∵ no filters)
- ▶ The advertisers incur less advertising costs (∵ higher response rate)

## Limitations?

- ▶ No incremental adoption
- ▶ Other Nash Equilibria?

# Conclusions

- ▶ There is not just a supply but a demand curve for UCM
- ▶ The open channel induces UCM suppliers and demanders to move out of the current email system
- ▶ It can be implemented with little cost, and is reversible if it doesn't work!

Thank you!

Your ideas please