

Using Uncensored Communication Channels to Divert Spam Traffic*

Benjamin Chiao[†] and Jeffrey MacKie-Mason[‡] 1/29/08 (First Version: 8/31/06)

Updated version: <http://benjaminchiao.org/papers/openchannel.pdf>

{`bchiao, jmm`}@umich.edu

Abstract

We offer a microeconomic model of the two-sided market for the dominant form of spam: bulk, unsolicited, and commercial advertising email. A key insight is our recognition that the bulk of spam is advertising, and to model the problem as a problem in the market supply and demand for advertising, rather than the usual approach of modeling spam as pure social cost to be eliminated for all. We adopt an incentive-centered design approach to develop a simple, feasible improvement to the current email system using an uncensored (open) communication channel. Such a channel could be an email folder or account, to which properly tagged commercial solicitations are routed without any blocking or filtering along the way. We characterize the circumstances under which spammers would voluntarily move much of

*We appreciate comments from Nat Bulkeley, Zhuoran Chen, Angus Chu, Nick Economides, Joseph Farrell, Michael Hess, Peter Honeyman, Paul Resnick, Doug Van Houweling, Russell Neuman, Hal Varian, Mike Wellman, the members of the Incentive-Centered Design Lab (especially Josh Cherry, Greg Gamette, Lian Jian, Kil-Sang Kim, John Lin, Anya Osepayshvili, Toinu Reeves, Ben Stearns, and Rick Wash) and the Michigan China Fellows (especially Ye Du, and Lei Zhong), and participants at the STIET workshop in May 2006, the Telecommunications Policy Research Conference, and the Fourth Bi-annual Conference on the Economics of the Software and Internet Industries. We gratefully acknowledge financial support from the NET Institute and from NSF grants IIS-0414710 and IGERT-0114368.

[†]School of Information, Univ. of Michigan, Ann Arbor, MI 48109

[‡]Dept. of Economics, and School of Information, Univ. of Michigan, Ann Arbor, MI 48109

their spam into the open channel, leaving the traditional email channel dominated by person-to-person, non-spam mail. Because there is real demand for unsolicited commercial email everyone can be made better off if a channel is provided for spammers to meet spam-demanders. As a bonus, the absence of filtering in an open channel restores to advertisers the incentive to make messages truthful, rather than to disguise them to avoid filters. We show that under certain conditions all email recipients are better off when an open channel is introduced. Only recipients wanting spam will use the open channel enjoying the less disguised messages and cheaper sale prices, and for all recipients the dissatisfaction associated with both undesirable mail received and desirable mail filtered out decreases.

1 Introduction

We all receive spam; we all resent it. Justice Potter Stewart, were he alive, would know it when he saw it. Nonetheless, it is hard to find a consensus definition of spam. Some want to include all unsolicited commercial email; others include unsolicited bulk email; others distinguish between deceptive, informative or malicious email. We should not be surprised, then, that it is also hard to find systematic analyses of “the spam problem”, when there are so many notions of what spam is.

Our goals are twofold. We want to identify a particular (but prevalent) subspecies of spam, analyze its ecology, and propose a mechanism that may increase social welfare substantially by modifying the flows of this type of spam. We hope to lay groundwork for systematic modeling of spam, and the consequent development of solutions that are effective because they address systematic features of the problem.

We limit our consideration to spam defined as *bulk, unsolicited, commercial email*; that is, effectively identical (but usually randomly disguised) messages sent unsolicited to large numbers of recipients with the goal of inducing a willing, mutually-beneficial purchase by the recipients. With this definition (we will call it “spam” for convenience, but it’s merely one subspecies) we rule out malicious bulk unsolicited email (e.g., email carrying a virus payload); we rule out deceptive email (e.g., “phishing” messages that attempt to trick recipients into revealing valuable personal information such as bank passwords); and we rule out email (though *initially* unsolicited) sent to a mailing list, from which one could unsubscribe. This is not to say that

what we rule out is insignificant; we just take our definition as a starting point and leave the study of other email problems for future research.

Defined as we have done, spam is an instance of a differently-named, well-known phenomenon: advertising. Using the less-pejorative moniker “email advertising” might give us a good start on a thoughtful, systematic consideration; certainly, it might help us recognize that at least this type of spam is not per se evil or morally deficient (though, as with any advertising, some population subgroups might conclude that the products advertised might fail that group’s morality test). Nonetheless, we will use “commercial spam” or just “spam” for short, because we relish the powerful affective response the term receives, and the opportunity to puncture the pejorative bubble it engenders.

To develop a systematic analysis of (non-deceptive, non-malicious) commercial spam, we need grounding principles. We find that surprising insights follow from adopting just two familiar, simple economic principles:

Revealed preference There is a non-trivial *demand* for the receipt of spam email.

Rational choice Spam purveyors will send spam messages to whomever, wherever, whenever, as long as the expected benefits exceed the expected costs.

We expect that only the first principle will raise many eyebrows at first, but we find that the second principle consistently has been half-ignored in most prior literature on “the spam problem”.

First, demand. Spam is not costless to generate or deliver, despite casual claims to the contrary. It is true that replication and transport costs are extremely low, compared to non-digital advertising channels. But there are a number of other costs: marketing and contracting costs with advertisers, content creation costs, content disguising costs (to get past technological filters), distribution technology costs (most spam is now sent out by virus-created spambots running on many machines not owned by the spam provider; these botnets need to be continuously regenerated, which requires developing new viruses to distribute, among other things). There may also be the cost of expected legal penalties. Given the non-zero costs of providing a spamming service, and with our scope limited to just commercial spam, from which the benefit to the sender is the inducement of willing purchases by recipients, we must conclude the following: by revealed preference,

there is a non-trivial demand for the receipt of spam email. Some consenting adults must be purchasing enough Rolex knock-offs and counterfeit products to pay the spammer’s costs. While the revealed demand could encompass some spurious demand induced by malicious or deceptive products (e.g. fake Viagra), a portion of the revealed demand is likely to be real. Few buyers will believe that a \$50 Rolex is authentic. We are not ruling out spurious demand completely. To the extent that recipients continue to believe in deceptive messages, they could find it beneficial to opt-in to the open channel to satisfy such spurious demand. Just another instance of placebo effects: inert imitation Viagra could work simply because people think it works.

To motivate and clarify our scope, we present some informal evidence. Such evidence is consistent with our claim that there is non-trivial demand for much spam: Cranor and LaMacchia [1998] show that the largest fraction of spam content is commercial advertising for products hard to find through other advertising channels. Sophos [2005] finds that this pattern continues; for example, in 2005 medication spam constitutes around 40% of all spam, and adult content for another 10-20%. Evett [2006] estimates that product spam constitutes around 25% of all spam, and adult content for another 19%.¹ We expect there will always be significant demand for “push” advertising in addition to “pull” (search-based) advertising.²

Recognizing that *some* recipients want to read spam, while many others evidently do not, we immediately see that one opportunity for social welfare improvement is to find a way to match commercial spam to those who want it, and not to those who do not. The latter email readers would benefit, and spam senders would also benefit by not incurring the costs of sending to people who will not want to purchase.

As a corollary, we expect the willing recipients of commercial spam to

¹Evett [2006] compiles his statistics from sources that include Google, Brightmail, Jupiter Research, eMarketer, Gartner, MailShell, Harris Interactive, and Ferris Research.

²This is evidenced by the multiple media for advertising that co-exist in equilibrium (Yellow Pages, local newspapers, billboards, broadcast TV and radio ads, bulk unsolicited commercial surface mail ads, etc.). Many products using commercial spam advertising do not want a durable, public presence. If they are moving their web sites to new domains frequently, they need a communication channel through which to disseminate each new, temporary location. Indeed, we observe cases in which the links for some domains selling medications expired in Google’s index well before Google got a chance to renew the links. MessageLabs [2005] shows that about 30% of spam domains expire within 24 hours. That’s why later in this section one could see that an open channel satisfies a market for push-advertising, as opposed to the pull-advertising of the World Wide Web.

benefit as well: if spammers can find a way to send to those who are interested in receiving the advertisements, then they can reduce their costs *and* increase the information content and quality in their ads, to the benefit of those who want the commercial information. Consider: Yellow Pages are a fairly successful bulk advertising medium because its ads are generally viewed only by those who want to see them, and the advertisers have the incentive to make the ads clear and informative, giving the viewers the information they desire. Spammers in contrast incur substantial costs to disguise the information in their ads so that filters cannot easily remove the ads from the email stream. But then the readers who do want the information so they can make a purchase are confronted with uninformative, low-value ads.

The second principle we offer as a foundation for systematic analysis of the spam ecology is that spammers are for the most part rational business-people, and they will send ads when the expected benefit to them exceeds the expected cost. What insight do we obtain from this unsurprising observation?³ We answer, first, indirectly: most other authors addressing spam have focused on proposals to raise the cost of spamming as a way of reducing the amount of spam produced. This approach is principled, but incomplete. An equivalent reduction in the benefits of spamming (e.g., by inducing those who want spam to read it in a different channel) should have the same (qualitative) incentive effect. If spam were flood waters, the existing solutions are in the spirit of building stronger levees to raise the river banks, instead of diverting the flood waters using a floodway. Both might properly belong in an effective flood management policy.

We build on these two principles to construct a model for commercial spam that includes advertisers, spam service providers, email service providers, and mail recipients who have heterogeneous tastes for receiving spam.⁴ See Figure 1. We then introduce a simple but novel mechanism motivated by the two principles above: an uncensored (open) communication channel through which spam will be accepted without filtering or other attempts to block. Such a channel could be as simple as a standardized mail client folder that

³We know, of course, that not every decision, in every circumstance, satisfies a test for decision-theoretic rationality. We only require that costly business decisions in general follow from reasonable comparisons of benefits to costs.

⁴In our current model we focus on the preferences and behaviors of recipients, spammers, and advertisers. We use a reduced-form, non-adaptive representation for email service providers.

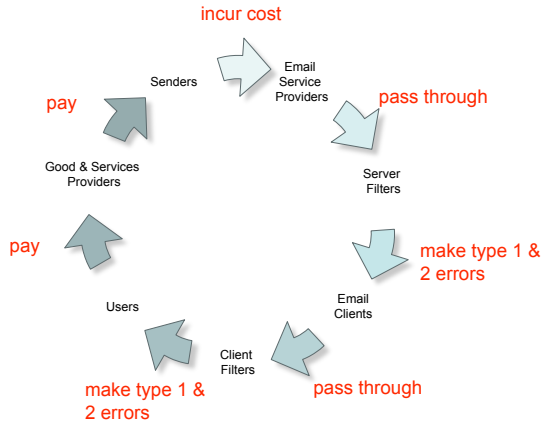


Figure 1: Stakeholders in an e-mail ecosystem.

would accept all appropriately labeled messages⁵. See Figure 2. Our conjecture is that if well-designed, then under some circumstances the introduction of an uncensored channel could result in substantial self-segregation by spammers, with email advertisements mostly targeted at the open channel, and much less at the traditional (censored) channel. See Figure 3.

There should be little dispute that if users could implicitly opt-in for commercial spam by creating an uncensored folder, the spammers would send mail to that channel. But why would they stop sending (or at least send less) to the censored channel? Our hypothesis is that if enough of the latent demand for purchasing spam-advertised products is reached through the uncensored channel, then the remaining commercial benefits obtainable from also spamming the censored channel may fall sufficiently low that they

⁵On a practical level, the sender chooses whether to send to the censored or the open channel (or both). When sending to the open channel, the sender does not disguise content, and adds a tag that indicates the message should be delivered to the open channel. If sending to the censored channel, the sender does not tag the message, and in fact may expend some effort to disguise the content.



Figure 2: An hypothetical open channel.

no longer justify the incremental costs.⁶ In our formal model below we show one set of conditions sufficient to guarantee this result.

There is another reason for spammers to keep sending to the traditional censored channel: persuasion. We are assuming that recipients know if they want to periodically purchase based on spam advertisements, and thus can make an ex ante rational choice about which channel to read. This situation is known in the literature as informative advertising⁷: consumers know they want information (price, location, etc.) about particular products, and seek out informative advertising to obtain the information they need. But there is another category: persuasive advertising, intended to convince consumers

⁶One reason the benefits in the censored channel might drop is that if users looking for unsolicited ads turn first to the better organized and more informative open channel, the level of demand for products advertised with less informative messages in the open channel may fall sufficiently that the expense of sending to the censored channel — and spending to get around the filters — is no longer worthwhile.

⁷See especially the section under the sub-heading “Is advertising used to inform or persuade?” on p. 28 of Taylor [1934].

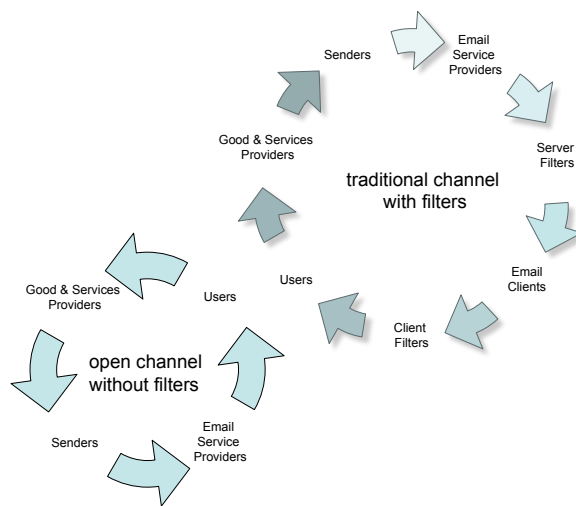


Figure 3: Separating the demand for and supply of bulk unsolicited commercial advertising.

to buy products they previously did not realize they wanted. Since these ads are aimed at consumers who might generally opt out of the open channel, it would do little good to send them to the open channel (which these previously uninterested customers shun), so the persuasive advertiser will generally go to where the unpersuaded are (the censored channel). Persuasion, however, comes at the cost of being filtered more easily because of the less disguised content. In any case, for this analysis we do not deal with the persuasive subspecies of unsolicited commercial email. If the open channel does not make persuasion in the censored channel *easier* than when there is only the censored channel, the omission of persuasive advertising does not bias our theoretical results.

Recall also that if spammers do choose to target the open channel, then we expect that they will also stop dissipating resources on unproductive efforts to disguise the informative content of their messages. Then those who wish to receive email advertisements will benefit from the higher quality (informativeness). This increase in informativeness, in turn, likely would induce a larger number of consumers to *want* to receive commercial spam.

We construct a model so that we may formally identify conditions under which the conjectures above hold true (and conditions under which they do not). Our main results are to characterize the degree to which spam will be shifted to the open channel, and to demonstrate that all parties benefit from the introduction of an open channel, so that it constitutes a Pareto improvement.

2 Prior approaches to spam

To date, most research focuses on reducing spam generally, usually through policy, technical or market mechanisms that raise the cost of sending spam. Before we detail our model of a mechanism that diverts spam to those who want it, and away from those who don't, we review other approaches.

2.1 Technological

Technological solutions have gained some partial success but the results are far from satisfactory even though they have been implemented for some time. The proposals include rule-based, Bayesian, and community (“collaborative”) filtering, disposable identities using extended email addresses [Ble-

ichenbacher et al., 1998], DomainKeys Identified Mail [Perez, 2005], Sender ID or Sender Policy Framework [Crocker, 2006]⁸, challenge-response [Dwork and Naor, 1993, Laurie and Clayton, 2004], whitelists, and blacklists. See Cranor and LaMacchia [1998] for an overview of these ideas (though not of course of the more recent specific proposals).

There is a fundamental problem with technological systems: they typically rely on the cost to spammers of devising technological workarounds. If the cost is high enough, the net benefit of spamming will be insufficient and the quantity of successful (delivered) spam will fall. However, the costs of technological workarounds fall rapidly, as technology becomes exponentially cheaper and as algorithmic solutions to hard computational problems rapidly improve. Thus, as the workaround cost falls, the technological barrier becomes less effective and spam delivered increases. This fundamental cost dynamic creates a need for ongoing investment to create improved anti-spam technologies. While an “arms race” may not be the first-best solution, we have not seen feasible methods to avoid this cycle, given the inevitable and rapid decline in technology costs. Certainly, though some of the cost has been shifted to ISPs and mail service providers, it appears that the social cost of spam has been increasing, not decreasing, despite the proliferation of technological fixes.

2.2 Legal

Legal rules are another approach to spam reduction. The U.S. CAN-SPAM act required a formal recommendation from the Federal Trade Commission regarding the establishment of a do-not-spam registry similar in the spirit of the do-not-call and do-not-fax registries created pursuant to the Telephone Consumer Protection Act of 1991. Although The FTC recommended against the creation of the list, other CAN-SPAM rules took effect 1 January 2004. However, legal solutions alone are, and likely will remain incomplete. First, to avoid prohibiting desirable email communications, legal rules generally include safe harbor provisions guaranteeing the permissibility of email exhibiting certain characteristics. It is generally difficult or impossible to prevent spammers from composing their messages so that they exhibit these characteristics, thus creating a safe harbor for a large and probably growing

⁸As of now, spam-sending domains are ironically the biggest users of SPF tags [MX-Logic, 2005]

quantity of spam. Second, legal jurisdiction over spam-distributing organizations is a crucial problem: spammers can easily change their locations to other countries.

2.3 Markets

Some proposals based on economic incentives have been gaining attention. These share an important feature with our approach to the problem: they typically are based on a presumption that users have heterogeneous values for receiving various email messages.

In an experimental investigation of email stamps as a price for obtaining a recipient's attention, Kraut et al. [2005] found that charging causes senders to be more selective and to send fewer messages. This method, however, requires non-spammers to pay a price as well. van Zandt [2004] examines the design of an optimal tax that minimizes exploitation of attention through information overload. Various email stamp systems have been or are about to be implemented.⁹ Loder et al. [2006] propose an attention-bond mechanism in which a sender deposits a monetary bond to a third-party agent, to be released only if the receiver tells the agent to do so. Both Loder et al. [2006] and we recognize heterogeneous valuations of mail messages; we particularly focus on those recipients who value the messages that are considered spam by most recipients.

Payment systems require substantial infrastructure for full implementation. The infrastructure necessary for widespread micropayment is lacking, and for successful adoption into a service exhibiting network effects, such as email, it is likely necessary that there be early widespread, not incremental, adoption, which is difficult to socially engineer. Also, there is a norm of free email service. Legitimate senders may resist paying for outgoing email more strenuously than is strictly justified if they took into account the system benefits to their recipients.

⁹Two of the world's largest providers of e-mail accounts, America Online and Yahoo!, announced in early 2006 that they would give preferential treatment to messages from companies paying from 1/4 of a cent to a penny each. An email stamp system was already implemented in Korea in 2003. Daum Corporation, the largest portal in Korea, charges about 0.8 cents to the senders who send more than 1000 messages per day. Fees scale downwards if senders are ranked lower than the biggest senders or more users rate the email as useful. Data cited by Kraut et al. [2005] indicate that spam was reduced by about 40% from its peak in a half-year period around the implementation.

3 Theory

In our brief review of other approaches to spam we highlighted one common feature: they are generally based on raising the costs of spamming, not on reducing the benefits. In addition, technological and legal methods (and some market methods, but less so) implicitly assume that certain mail (or mail senders) are uniformly undesirable; that is, they ignore heterogeneity in recipient preferences. In this section we present a model of the two-sided market for commercial spam, in which product sellers pay spammers to deliver advertisements to email recipients, some of whom in turn willingly choose to purchase the advertised products.¹⁰ We then analyze the effect of introducing an open (i.e., uncensored) channel. The open channel approach is designed to lower the benefits to spammers of sending mail to all recipients, and works only and precisely because recipient preferences are heterogeneous: viz., some recipients want to receive email advertisements.

3.1 Mail Types

Mail types coincide with senders' types. Such types are defined by two attributes: mass or targeted mail, and solicited or unsolicited.

The first attribute is mainly a cost attribute of sending. The content creation cost per copy of mass mail is much lower than that of targeted mail.¹¹ Also, because of information asymmetry of each recipient's preference for spam, by definition, mass-mail senders' best strategy is to randomize recipients' addresses.

The second attribute is mainly a cost attribute of blocking. A mail is considered solicited for our purpose even if it is initially unsolicited if one could easily unsubscribe (block) oneself from such mailing lists permanently.

In all, we identify four types of mail, and provide examples:

Unsolicited mass: Viagra and erotic content advertisements.

¹⁰The email market is a typical problem of two-sided markets (e.g., Rochet and Tirole [2003], Parker and Alstyn [2005]), which is closely related to the chicken-and-egg problem. Roughly speaking, the number of senders affects the number of recipients, and vice versa.

¹¹We do not require that it is possible to identify whether a message is mass mail or targeted mail. It is easy to fool general purpose filters, and the recipient often will not know until after incurring the cost of viewing the message.

Unsolicited targeted: Personalized advertisements based on purchase history obtained elsewhere.

Solicited mass: Advertisements from conventional booksellers, non-profit fundraisers, and other legal and less socially objectionable purveyors.

Solicited targeted: Personal correspondence.

Unsolicited mass mail constitutes the bulk of the unwanted email for most individuals. We therefore simplify our analysis using this convenient assumption:

Assumption 1 *Mail Segregation: Mass-mail senders send only unsolicited mail, and targeted-mail senders send only solicited mail.*

Henceforth, when the context is clear, we refer to unsolicited mass mail senders as senders, and the mail they send simply as mail.

3.2 The Recipients' Problem

Assume that for unsolicited mass mail, some individuals want to receive a fraction $1 - \epsilon$. $\epsilon \in [0, 1]$ increases if there are more undesirable properties with unsolicited mass mail such as phishing and malicious content. There is a large number of high-type (\bar{r}) recipients indexed by \bar{r} on the interval $[0, 1]$. The corresponding index for low-type recipients is \underline{r} . Only high-types desire unsolicited mass mail (that is, have some demand for the goods advertised in such mail); such preference is denoted by w_r , where $w_{\bar{r}} = 1$ and $w_{\underline{r}} = 0$. Recipients are willing to spend up to w_r on goods. They buy goods from ads in the channel that offer the lower average price, where p^j is the average price offered by ads in channel j . We will later see that they respond to the advertisement based on how informative or transparent the mail is. Whether mail (desired or undesired) is received depends on the filtering technology employed by the email service provider. We model this below, but for now simply refer to mail that gets through as “unfiltered” and mail that does not as “filtered”.

In the censored channel, filtering technology is designed to block unsolicited mass mail, but it does so imperfectly. Each sender knows that the filter has a strength of $f^c \in [1, \infty)$ for unsolicited mass-mail. The filter

strength is simply the inverse of the fraction of mail that gets through the filter. By definition there is no filtering in the open channel, $f^o = 1$.¹²

Sender s can make an effort to disguise its content to reduce the filter's success rate. We let sender s choose a transparency level, $t_s^j \in [\frac{1}{f^j}, 1]$, for mail sent to channel j , where t_s^j is a multiplicative factor adjusting the filter strength. If $t_s^j = 1$, the mail is transparent with no disguise at all and the effective filter strength is the technological strength f^j . If $t_s^j = 1/f^j$, the effective filter strength is one, which is to say, all content passes through unfiltered. Disguising is costly; there is no effort made to disguise content in the open channel.¹³

Denote N_r^j as the volume of mail sent to recipient r in channel j . Essentially, it is the mail sent averaged across all recipients in j . R_r^j is the number of high-types using channel j ; R^j is the total number of recipients who choose to use channel j . Then the unfiltered portion that actually reaches recipient r is denoted by the small letter $n_r^j \equiv \frac{N_r^j}{t^j f^j}$, where t^j is the weighted average of transparency levels in channel j .

We define $\kappa_r^j = 1$ if recipient r uses channel j , zero otherwise. To build a tractable model, we make another assumption to rule out the *unlikely* scenario that no one is using the existing email channel:

Assumption 2 Channel Essentiality: *The censored channel is essential so that every recipient uses it. That is, $\kappa_r^c = 1$.*

Let us now state the recipient's problem formally. Given other variables, recipient r makes a binary choice of whether to opt into the open channel, $\kappa_r^o \in \{0, 1\}$, by maximizing:

$$U_r(\kappa_r^o) \equiv U_r(v_r^g(\kappa_r^o), v_r^I(\kappa_r^o), v_r^{II}(\kappa_r^o)) \quad (1)$$

U^r is increasing in the first argument, and decreasing for the rest. The first argument is the volume of goods consumed, which is given by total spending divided by price (paid to the advertiser indexed by a) in the channel with

¹²An approach, which is perhaps less radical in practice, is to extend the current model such that $1 < f^o < f^c$. Our central results should still hold for f^o to be sufficiently small. The magnitude of f^o , however, is an empirical question. We therefore only solve for the baseline case of $f^o = 1$, which should have the same qualitative effects.

¹³By definition of the lower bound of t_s^j , $t_s^o = 1$ because $f^o = 1$ implies that the upper and lower bounds coincide.

the lower price (and which is subscribed to by the recipient):

$$v_r^g = \max_j \left\{ \frac{\kappa_r^j w_r}{p_a^j} \right\}^{14}, \quad (2)$$

which is zero for low-type recipients since $w_r = 0$.

The second argument gives the Type I errors (unwanted mail that is received):

$$v_r^I = \sum_{j \in \{o, c\}} (1 - w_r + w_r \epsilon) \kappa_r^j \frac{N_r^j}{t^j f^j}. \quad (3)$$

When a recipient is a high type ($w_r = 1$), this is the fraction ϵ of unsolicited mail the high type does not want to receive (summed across the channels to which she subscribes). For a low type, this is all unsolicited mail received.

The third argument gives the Type II errors (wanted mail that is filtered out before delivery):

$$v_r^{II} = \sum_{j \in \{o, c\}} (1 - \epsilon) w_r \kappa_r^j N_r^j \left(1 - \frac{1}{t^j f^j}\right), \quad (4)$$

which is zero for low-type recipients since $w_r = 0$. There is no filtering in the open channel, so for high types this is the fraction of desirable unsolicited mail ($1 - \epsilon$) that is filtered out of the censored channel.

For high-type recipients, $\frac{\partial U_r}{\partial v_r^g} \rightarrow \infty$ as $v_r^g \rightarrow 0$ and $\frac{\partial U_r}{\partial v_r^g} < \infty$ as $v_r^g \rightarrow \infty$. For all recipients, $\frac{\partial U_r}{\partial v_r^I}, \frac{\partial U_r}{\partial v_r^{II}} > -\infty$ as $v_r^I, v_r^{II} \rightarrow \infty$. These are the Inada conditions commonly used to denote factors of production that are essential or bounded in some sense.

3.3 The Senders' Problem

The total cost function for sender s is $C_s(N_s^o, N_s^c, t_s^c) \equiv c_s^o N_s^o + c_s^c N_s^c + d(t_s^c)$. It captures the costs of generating the email volumes, and of reducing the transparency of the mail:

¹⁴We simplified the problem from an equivalent but more explicit formulation: Recipient r chooses $\kappa_r^o \in \{0, 1\}$ and $\sum_j v_r^{g,j}$ to maximize $U_r(\kappa_r^o, \sum_j v_r^{g,j} | \cdot) \equiv U_r(\sum_j v_r^{g,j}, v_r^I, v_r^{II})$ s.t. $\sum_j \kappa_r^j p_a^j v_r^{g,j} = w_r$.

- constant returns to scale in each volume cost: $\frac{\partial C_s}{\partial N_s^c} = c_s^c > \frac{\partial C_s}{\partial N_s^o} = c_s^o > 0$ ¹⁵. Note that c_s^c and c_s^o are constants.
- negative transparency (positive disguise) cost: $\frac{\partial C_s}{\partial t_s^c} < 0$

On the revenue side, senders are price takers. Advertisers pay them for solicitations. Let p_s^j be the advertising charge per disguised email *reaching* the users in channel j ¹⁶.

Given other variables, sender s chooses (N_s^o, N_s^c, t_s^c) to maximize:

$$\pi_s(N_s^o, N_s^c, t_s^c) = p_s^o N_s^o + \frac{p_s^c N_s^c}{t_s^c f^c} - c_s^o N_s^o - c_s^c N_s^c - d(t_s^c) \quad (5)$$

$$\text{s.t. } t_s^c \in \left[\frac{1}{f^c}, 1\right], N_s^o, N_s^c \geq 0 \quad (6)$$

Next we state the solutions to the above maximization problem:

Result 1 *The best responses of sender s are:*

$$N_s^o > 0 \iff p_s^o \geq c_s^o \quad (7)$$

$$N_s^c > 0 \iff \frac{p_s^c}{t_s^c f^c} \geq c_s^c \quad (8)$$

$$t_s^c \begin{cases} = 1 \\ \in \left(\frac{1}{f^c}, 1\right) \\ = \frac{1}{f^c} \end{cases} \iff N_s^c c_s^c \begin{cases} < \\ = \\ > \end{cases} - t_s^c \frac{\partial C_s}{\partial t_s^c} \quad (9)$$

¹⁵Rather than having a zero marginal sending cost ($\partial C_s / \partial N_s^j$) as commonly asserted, spammers incur cost to renew technologies, which depreciate quickly, to generate spam. For example, zombies (i.e., home computers hijacked by crackers) are consistently destroyed by anti-virus software, so spammers must continuously develop and distribute new viruses to capture new (temporary) zombies. Zombies are responsible for relaying more than 60% of the world's spam [Sophos, 2005]. One could have used a step function to model the cost function but a smooth function would approximate a step function when the number of zombies increases.

¹⁶In practice, there is a volume discount (that might, for instance, be due to diminishing likelihood to respond). For instance, Send-Safe is a service spammers offer to advertisers. One pricing scheme asks for US\$125 per 1 million credits (possibly a proxy of $\frac{N_s^c}{t_s^c f^c}$) when an advertiser pays for 0.4 million credits. The price drops monotonically to US\$10 per 1 million credits when an advertiser pays for 300 million credits. This pricing scheme is available at <http://www.send-safe.com/send-safe.html>, accessed November 15, 2006.

Proof. See Appendix 6.1. ■

Notice that the marginal revenues of sending N_s^o and N_s^c are p_s^o and $\frac{p_s^c}{t_s^c f^c}$. When the marginal revenue of channel j is strictly less than the marginal cost, sender s does not send any to channel j . Else, sender s makes some economic profit because in channel j the marginal revenue is weakly greater than the marginal cost in a given channel.

3.4 The Advertisers' Problem

There is a large number of advertisers indexed by a on the interval $[0, 1]$. Advertiser a 's total cost derives from the production of goods sold, and from the advertising for them.

Let θ^j be the probability that an advertisement in channel j leads to a purchase, and n_a^j is the number of messages delivered on behalf of advertiser a (whereas N_a^j is the number sent, the difference being due to filtering). Production cost $C_a(\sum_j \theta^j n_a^j) \equiv c_a \cdot \sum_j \theta^j n_a^j$, exhibits constant-returns-to-scale technology. c_a is the constant marginal production cost. $\theta^j = \theta(t_a^j, \frac{R_r^j}{R^j})$, where t_a^j is the transparency level of email in channel j associated with advertiser a , and $\frac{R_r^j}{R^j}$ is the ratio of high-type recipients in channel j . θ^j is increasing in both arguments; it equals zero if $\frac{R_r^j}{R^j} = 0$.

Advertising cost, $\sum_j p_s^j n_a^j$, is linear. Recall that p_s^j is the marginal advertising charge in channel j that is paid to senders.

For each p_s^j paid, advertiser a collects expected sales revenue equal to the price paid by ad recipients *cum* consumers, p_a^j , times the probability of making a sale in channel j , θ^j .

Given other variables, advertiser a chooses (n_a^o, n_a^c) to maximize:

$$\pi_a(n_a^o, n_a^c) = \sum_j (p_a^j \theta^j(t_a^j, \frac{R_r^j}{R^j}) - p_s^j) n_a^j - c_a \cdot \sum_j \theta^j n_a^j \quad (10)$$

$$\text{s.t. } n_a^j \geq 0 \quad (11)$$

Result 2 *The best responses of advertiser a are:*

$$n_a^c > 0 \iff \theta^c c_a + p_s^c \leq \theta^c p_a^c \quad (12)$$

$$n_a^o > 0 \iff \theta^o c_a + p_s^o \leq \theta^o p_a^o \quad (13)$$

Proof. See Appendix 6.2. ■

This result implies that the volume requested should be zero in both channels if the marginal cost exceeds the marginal benefit in each channel, and that the volume requested in channel j should be strictly positive when the marginal markup is positive. The marginal markup is the marginal sale price (p_a^j) minus the marginal cost of a sure response ($\frac{p_s^j}{\theta^j}$) and marginal production cost (c_a).

3.5 Equilibrium

3.5.1 Competitive Equilibrium

By assuming atomistic, price-taking recipients, advertisers and spam senders (which can be justified for the latter two by an assumption of free entry), we have the conditions for a competitive equilibrium. Using a hat symbol to denote equilibrium values, a competitive equilibrium by definition satisfies the following:

- the goods and services market is cleared:

$$\int_{\bar{r} \in [0,1]} \hat{v}_{\bar{r}}^g d\bar{r} + \int_{r \in [0,1]} \hat{v}_r^g dr = \sum_j \hat{\theta}^j \left(\int_{\bar{r} \in [0,1]} \hat{n}_{\bar{r}}^j d\bar{r} + \int_{r \in [0,1]} \hat{n}_r^j dr \right) \quad (14)$$

- the mail market is cleared:

$$\int_{\bar{r} \in [0,1]} \hat{n}_{\bar{r}}^j d\bar{r} + \int_{r \in [0,1]} \hat{n}_r^j dr = \int_{a \in [0,1]} \hat{n}_a^j da = \int_{s \in [0,1]} \frac{\hat{N}_s^j}{\hat{t}_s^j f^j} ds \quad (15)$$

- the profits of senders and advertisers are maximized:

$$\pi_s(\hat{N}_s^o, \hat{N}_s^c, \hat{t}_s^c) = \max_{N_s^o, N_s^c, t_s^c} \pi_s(N_s^o, N_s^c, t_s^c) \quad (16)$$

$$\pi_a(\hat{n}_a^o, \hat{n}_a^c) = \max_{n_a^o, n_a^c} \pi_a(n_a^o, n_a^c) \quad (17)$$

- the zero economic profit conditions are satisfied:

$$\pi_s(\hat{N}_s^o, \hat{N}_s^c, \hat{t}_s^c) = 0 \quad (18)$$

$$\pi_a(\hat{n}_a^o, \hat{n}_a^c) = 0 \quad (19)$$

- the utility of each of the high and low-type recipients is maximized:

$$U_{\bar{r}}(\hat{\kappa}_{\bar{r}}^o) = \max_{\kappa_{\bar{r}}^o} U_{\bar{r}}(\kappa_{\bar{r}}^o) \quad (20)$$

$$U_{\underline{r}}(\hat{\kappa}_{\underline{r}}^o) = \max_{\kappa_{\underline{r}}^o} U_{\underline{r}}(\kappa_{\underline{r}}^o) \quad (21)$$

3.5.2 Aggregations

By homogeneity, integrating over all agents of the same type gives a commonly used result: the aggregate value equals to the average value (i.e. $\int_{k \in [0,1]} \hat{z}_k^j dk = \hat{z}_k^j$). In this sense, we justified the use of representative agents at least for its convenience. The values they have chosen are simply the total consumed by the same type of agents in the economy.

The following expression states that the volume of goods consumed by all recipients (and the representative high and low type recipients) equals to the response rate times the unfiltered volume received, requested or sent across all channels. (14) and (15) give:

$$\sum_j (\hat{v}_{\bar{r}}^{g,j} + \hat{v}_{\underline{r}}^{g,j}) = \sum_j \hat{\theta}^j (\hat{n}_{\bar{r}}^j + \hat{n}_{\underline{r}}^j) = \sum_j \hat{\theta}^j \hat{n}_a^j = \sum_j \hat{\theta}^j \frac{\hat{N}_s^j}{\hat{t}_s^j f^j}. \quad (22)$$

3.5.3 Nash Equilibrium

We first show the following best response of the recipients. High type recipients, and only they, will use the open channel if the mail volume sent to the open channel is strictly positive and that sent to the censored channel approaches zero:

Result 3 *If $N_r^o > 0$, $N_r^c \rightarrow 0$, then $\kappa_{\bar{r}}^o = 1$ and $\kappa_{\underline{r}}^o = 0$ is the best response.*

Proof. If $\kappa_{\bar{r}}^o = 1$, $v_{\bar{r}}^g$ strictly increases for high-type recipients because of the max operator in (2). There is no effect on Type II errors (4) because for the open channel $t^j = f^j = 1$, so the term for $j = o$ is zero. For low-type recipients, $v_{\bar{r}}^g$ and $v_{\bar{r}}^{II}$ are zero anyway, but $v_{\bar{r}}^I$ can only strictly increase if $\kappa_{\bar{r}}^c = 1$ and $N_r^o > 0$. For high-type recipients, the question then is whether the utility increase with the increase in $v_{\bar{r}}^g$ would at least offset the utility decrease with the increase in $v_{\bar{r}}^I$. Note that $N_r^c \rightarrow 0$ and $\kappa_{\bar{r}}^o = 0$ imply $v_{\bar{r}}^g \rightarrow 0$ (by (22)) and $\frac{\partial U_{\bar{r}}}{\partial v_{\bar{r}}^g} \rightarrow \infty$. But $\frac{\partial U_{\bar{r}}}{\partial v_{\bar{r}}^I} > -\infty$ when $v_{\bar{r}}^I \rightarrow \infty$. We must have $U_{\bar{r}}(\kappa_{\bar{r}}^o = 1) - U_{\bar{r}}(\kappa_{\bar{r}}^o = 0) > 0$. ■

We are now ready to show a Nash equilibrium in which (i) the open channel diminishes the volume of unsolicited mass mail in the censored channel to zero, and (ii) only high-type recipients will use the open channel.

Proposition 1 *Nash*: ($\hat{\kappa}_{\bar{r}}^o = \hat{\kappa}_{\bar{r}}^c = \hat{\kappa}_{\underline{r}}^c = 1, \hat{\kappa}_{\underline{r}}^o = 0; \hat{N}_s^o > 0, \hat{N}_s^c \rightarrow 0, \hat{t}^c \leq 1; \hat{n}_a^o > 0, \hat{n}_a^c \rightarrow 0$) is *Nash*.

Proof. Note first that $\hat{N}_s^o = \hat{n}_a^o$ and $\frac{\hat{N}_s^c}{\hat{t}_s^c f^c} = \hat{n}_a^c$. Suppose $\hat{N}_s^o > 0, \hat{N}_s^c \rightarrow 0, \hat{t}_s^c \leq 1, \hat{n}_a^o > 0, \hat{n}_a^c \rightarrow 0$, we need to show that $\hat{\kappa}_{\bar{r}}^o = \hat{\kappa}_{\bar{r}}^c = \hat{\kappa}_{\underline{r}}^c = 1$ and $\hat{\kappa}_{\underline{r}}^o = 0$. Assumption 2 implies $\hat{\kappa}_{\bar{r}}^c = \hat{\kappa}_{\underline{r}}^c = 1$. Result 3 implies $\hat{\kappa}_{\bar{r}}^o = 1$ and $\hat{\kappa}_{\underline{r}}^o = 0$ because $\hat{N}_s^o > 0, \hat{N}_s^c \rightarrow 0$.

Now suppose $\hat{\kappa}_{\bar{r}}^o = \hat{\kappa}_{\bar{r}}^c = \hat{\kappa}_{\underline{r}}^c = 1$ and $\hat{\kappa}_{\underline{r}}^o = 0$ (implying $\hat{\theta}^o > \hat{\theta}^c = 0$ since $\hat{t}_s^c \leq \hat{t}_s^o = 1$), we need to show that $\hat{N}_s^o > 0, \hat{N}_s^c \rightarrow 0, \hat{t}_s^c \leq 1, \hat{n}_a^o > 0$ and $\hat{n}_a^c \rightarrow 0$.

To show $\hat{n}_a^c \rightarrow 0$, note that the first-order condition in the advertiser's problem that determines \hat{p}_a^c is $\hat{\theta}^c c_a + \hat{p}_s^c - \mu^c = \hat{\theta}^c \hat{p}_a^c$, where μ^c is the Lagrangian multiplier satisfying $\mu^c \hat{n}_a^c = 0$. $\hat{\theta}^c = 0 \implies \mu^c > 0 \implies \hat{n}_a^c = 0 \implies \hat{n}_a^c \rightarrow 0$.

To show $\hat{n}_a^o > 0$, suppose otherwise that $\hat{n}_a^o = 0$. Together with $\hat{n}_a^c \rightarrow 0$ that we already proved, we have $\hat{v}_r^g \rightarrow 0$ by (22). $\hat{v}_r^g \rightarrow 0 \implies \hat{p}_a^o, \hat{p}_a^c \rightarrow \infty$ because $\hat{v}_r^g = \max_j \{ \frac{\kappa_r^j w_r}{\hat{p}_a^j} \}$ (from 2), which contradicts with $\hat{n}_a^o = 0$ because \hat{n}_a^o is not well defined. This is because with $\hat{\theta}^o > \hat{\theta}^c = 0, p_s^o \leq c_s^o < \infty$ (by Result 1), $c_a < \infty$, the objective function of the advertiser's problem (10) becomes $(p_a^o \theta^o - p_s^o) n_a^o - c_a \cdot \theta^o n_a^o = \infty$. n_a^o cannot be determined because this maximization problem is not well defined. ■

3.6 Welfare

When the sales price in all transactions is lowered, the volume of goods sold increases given a fixed expenditure. Equation (22) tells us it could mean that the response rate, the mail volume or both has increased. It could also mean that the response rate becomes so high that the mail volume decreases. Denote the status-quo variables when the open channel is absent using ∞ ¹⁷, we show the latter in Proposition 2.

(23) says that the mail volume received by *all* recipients is lower than that in the status quo if and only if the total marginal cost *per mail received* (i.e.,

¹⁷When the open channel is absent, we interpret it as $f^o \rightarrow \infty$. When such filter strength is infinitely strong, it is as if there is no such channel for any practical use.

the sum of marginal costs of production ($c_a \hat{\theta}^o$) and advertising (\hat{p}_s^o) per mail received) is higher than the status quo value ($c_a \hat{\theta}^{c,\infty} + \hat{p}_s^{c,\infty}$). For the mail volume to decrease after the implementation of the open channel, the new response rate has to be so high that the increase in the marginal cost of production per mail received ($c_a \hat{\theta}^o - c_a \hat{\theta}^{c,\infty}$) more than offsets the decrease in the marginal cost of advertising per mail received ($\hat{p}_s^{c,\infty} - \hat{p}_s^o$).

Since $\frac{R}{R_r} > 1$, (24) implies (23). When each recipient receives less, the total received is also less (the converse is not true). It is scaled by $\frac{R}{R_r}$ because the comparison now is between the total marginal costs per mail received by the *high-type recipients* (trivially the low-type recipients receive less because there is no unsolicited mass mail in the only channel they opted in). In the status quo, for every $\frac{R}{R_r}$ mail received by all recipients, only one is received by the high-type recipients, that is why the corresponding total marginal cost is inflated by $\frac{R}{R_r}$.

Proposition 2 Mail Volume: *For the Nash equilibrium in Proposition 1,*

$$\sum_j \frac{\hat{N}_{\bar{r}}^j + \hat{N}_r^j}{\hat{t}^j f^j} \leq \frac{\hat{N}_{\bar{r}}^{c,\infty} + \hat{N}_r^{c,\infty}}{\hat{t}^{c,\infty} f^c} \iff c_a \hat{\theta}^o + \hat{p}_s^o \geq c_a \hat{\theta}^{c,\infty} + \hat{p}_s^{c,\infty}, \quad (23)$$

$$\frac{\hat{N}_r^o}{\hat{t}^o f^o} \leq \frac{\hat{N}_r^{c,\infty}}{\hat{t}^{c,\infty} f^c} \iff c_a \hat{\theta}^o + \hat{p}_s^o \geq \frac{R}{R_r} (c_a \hat{\theta}^{c,\infty} + \hat{p}_s^{c,\infty} \hat{\theta}^o), \quad (24)$$

where $\hat{p}_s^{c,\infty} = c_s^c \hat{t}^{c,\infty} f^c > \hat{p}_s^o = c_s^o$.

Proof. See Appendix 6.3. ■

In the following proposition, we prove that in a Nash equilibrium described in Proposition 1, each recipient's utility has not decreased after the roll-out of the open channel because the utility associated with goods does not decrease, and the dis-utilities associated with Type II errors do not increase. Type I errors will not increase for high-type recipients if Proposition 2 holds or $\epsilon = 0$. Requiring free disposal of unsolicited mass mail in the open channel ($\epsilon = 0$) is justified in the sense that such mail messages, which advertise similar products, are highly substitutable if there is a good search engine within the open channel. Instead of $\epsilon = 0$, we also prove the case in which we require a lower unsolicited mass mail volume than in the status quo.

The structure of the proof is simple. The welfare of the unsolicited mass-mail senders and advertisers will be unchanged because they make zero profit

with and without the open channel. The welfare of the recipients could differ. To compare the welfare change for a given recipient when there is an open channel, we compare his or her utility after and before the roll-out of the open channel: $\hat{U}_r - \lim_{f^o \rightarrow \infty} \hat{U}_r$.

Proposition 3 Welfare: *If Proposition 1 holds and either Proposition 2 holds or $\epsilon = 0$, the welfare of the advertisers, unsolicited mass-mail senders and all recipients will be weakly increased when there is an open channel.*

Proof. See Appendix 6.4. ■

We leave it for further research to analyze the possibility of other Nash equilibria, and (if any) their characteristics.

4 Discussion

We emphasize that our proposal is a starting point. There are many issues beyond the scope of this analysis worth further investigation. We discuss a few here.

It may seem that an open channel is already implemented by the common mail client spam folder. However, a client spam folder is usually the last stage of multi-stage filtering. Most email service providers block some traffic altogether, and only mail that makes it through the first (or first $n - 1$) filter(s) are candidates for quarantine in a user spam folder. Thus, to get to that channel, senders already incurred disguise costs.

We mentioned at the beginning that an open channel is low-cost to implement. Any email service provider could immediately, at minimal cost, offer to let all mail to certain clients pass through with a tag in a header (or even pre-pended to the subject field) that would allow immediate and complete channeling to a separate mailbox; when the email service provider itself provides the mail client (as do online email providers like Gmail and Yahoo!) the open channel inbox could be provided for users that wanted. What is less clear is whether incremental implementation would be successful at drawing senders and readers of unsolicited commercial email into the open channel, or whether widespread, coordinated adoption would be necessary.

Would the open channel be flooded with so much more mail — now that the cost sending ads is lower — that those who want to see ads will incur costs so high to find desired ads that they abandon using the open channel? This is a complicated question, and one largely beyond the scope of this paper. There

are good reasons we think this may not happen, however. Since advertisers no longer need to disguise their messages to get into the inbox, they need to only send one message rather than many to guarantee delivery. Further, the advantages of having many more or less identical messages in the inbox may be reduced if recipients can easily find the messages they want and use indexing and finding services to sort through them. As an analogy, we are not aware of any advertisers that pay to place more than one display ad in a single yellow pages book. Another possibility is the implementation of an expiration feature to the open channel inbox: any messages beyond an expire date are automatically deleted, to reduce clutter. Of course, advertisers will still want to compete with each other for attention, and it is possible they will try to do so by proliferating messages in the open channel. The result on balance cannot be predicted without a more detailed model, and, most likely, some empirical evidence.

We have largely ignored what we call persuasive advertising, namely, advertising that tries to persuade people who did not know that they wanted to purchase anything. These recipients in general will not opt-in to the open channel, and so spammers may still try to reach them through the censored channel. Our conjecture is that if enough demand is shifted to the open channel, and especially if prices for goods advertised there fall due to the lower marketing costs, that demand in the censored channel for goods sold by persuasive spammers may fall enough to discourage this type of advertising, but we have not formally analyzed this.

5 Conclusions

Technical filters and legal rules raise the cost of delivering spam to readers. Costs are borne by spammers (who must develop ever-changing techniques for avoiding filters, etc.), but also by recipients, who spend time doing the difficult filtering and reviewing that cannot be automated, and paying higher costs for goods to cover the marketing expenses. On the other hand, an equivalent reduction in the benefits of spamming (e.g. by moving out spam demanders) should have the same incentive effect. More generally, methods that channel communications more directly to those who want them would lower costs on both sides and be welfare improving.

We formalized this intuition and explored sufficient conditions for all email recipients to be better off with the introduction of an open channel. We

show that under these conditions only recipients wanting unsolicited commercial advertisements will use the open channel, and they will benefit from less disguised messages and lower sales prices. In addition, for all recipients the dissatisfaction associated with both undesirable mail received and desirable mail filtered out decreases.

We do not claim that our idea would provide a complete solution to the current spam problem, but we do offer a novel new tool that, together with other well-known tools (technical, legal and economic), may contribute to a reduction in the flow of low-information, unsolicited commercial bulk email. The ultimate solution, simple economics predicts, is for the value of purchasing stimulated by spam to fall sufficiently low that it is less than the already low cost of sending spam. If we can tempt a substantial number of consumers who *want* to purchase spam-advertised products into a separate email channel, the purchasing value remaining in the traditional, filtered, or censored channel may drop sufficiently to discourage spammers from using that increasingly unproductive channel.

In other words, we take a straightforward economic approach to the question, by recognizing that there is not just a supply curve but also a demand curve for spam. We model the incentives, within the ecosystem of existing spam solutions, to induce both suppliers and demanders to move out of the current censored channel and into the open channel. If customers who want to purchase will benefit from more informative ads in a separate channel, then spam advertisers will benefit from focusing their advertising spending on that channel. This should not be a very controversial idea, but it is, we believe, an idea that has been largely missing from the debate.

There is another illuminating economic perspective on our work: spam is fundamentally a problem that arises when disposal is not free. We know from the First Fundamental Welfare Theorem that unregulated free markets are generally Pareto efficient, but that result requires free disposal. Spam is not free to dispose: it requires time to open and consider. Some types of spam are malicious and may actually cause harm to one's data files or operating system before we can dispose of it.

Our proposal recreates a free market — the open channel — for those who do not want to dispose of spam. It contrasts with other free-market solutions (e.g. email stamps and bonds for email spam, and Google's AdWords for web spam) in the following way. The open channel gives recipients the right to receive spam; it removes the right of the email service providers to decide whether the recipients should receive spam. (More generally, the recipients'

right to choose the level of censorship is one of the many other possible property right reassignments in the email ecosystem that have been largely unexplored in the literature.) Also, we provide those for whom the disposal costs are sufficiently high (not free) the choice to opt out and participate only in the censored channel. Meanwhile, senders (and spam demanders) do not internalize the disposal costs of uninterested recipients, but the senders nonetheless choose to send less to the censored channel because the average propensity to buy falls as spam demanders move to the open channel.

An open advertising channel is possible at low cost, and it is conceivable that it would make email users at least weakly better off (no worse off) than the status quo. At the very least, this mechanism is fully reversible. If well-designed, an incentive-compatible advertising channel that harnesses the simultaneous forces of demand and supply could significantly reduce the flow of unsolicited bulk commercial email.

6 Appendix

6.1 Proof of Result 1

The sender's profit function is

$$\pi_s(N_s^o, N_s^c, t_s^c) = p_s^o N_s^o + \frac{p_s^c N_s^c}{t_s^c f^c} - C_s(N_s^o, N_s^c, t_s^c) \quad (25)$$

The Lagrangian is:

$$\mathcal{L} = \pi(\cdot) - \lambda_1^c(t_s^c - 1) + \lambda_2^c(t_s^c - \frac{1}{f^c}) + \mu^o N_s^o + \mu^c N_s^c, \quad (26)$$

where $\lambda_1^c, \lambda_2^c, \mu^c, \mu^o \geq 0$.

The complementary slackness conditions are:

$$\lambda_1^c(t_s^c - 1) = 0 \quad (27)$$

$$\lambda_2^c(t_s^c - \frac{1}{f^c}) = 0 \quad (28)$$

$$\mu^o N_s^o = 0 \quad (29)$$

$$\mu^c N_s^c = 0 \quad (30)$$

FOCs:

$$p_s^o = \frac{\partial C_s}{\partial N_s^o} - \mu^o \quad (31)$$

$$\frac{p_s^c}{t_s^c f^c} = \frac{\partial C_s}{\partial N_s^c} - \mu^c \quad (32)$$

$$\frac{-p_s^c N_s^c}{(t_s^c)^2 f^c} - \lambda_1^c + \lambda_2^c = \frac{\partial C_s}{\partial t_s^c} \quad (33)$$

Case 1: $N_s^o, N_s^c > 0 \implies \mu^o = \mu^c = 0$.

(31) implies $p_s^o = \frac{\partial C_s}{\partial N_s^o}$. (32) implies $\frac{p_s^c}{t_s^c f^c} = \frac{\partial C_s}{\partial N_s^c}$.

Subcase 1: $t_s^c = 1 \implies \lambda_2^c = 0$

Combining (32) and (33), we have $-N_s^c \frac{\partial C_s}{\partial N_s^c} f^c - \lambda_1^c = \frac{\partial C_s}{\partial t_s^c}$. This implies $-N_s^c \frac{\partial C_s}{\partial N_s^c} f^c \geq \frac{\partial C_s}{\partial t_s^c}$.

Subcase 2: $t_s^c = \frac{1}{f^c} \implies \lambda_1^c = 0$

Combining (32) and (33), we have $-N_s^c \frac{\partial C_s}{\partial N_s^c} + \lambda_2^c = \frac{\partial C_s}{\partial t_s^c}$. This implies $-N_s^c \frac{\partial C_s}{\partial N_s^c} \leq \frac{\partial C_s}{\partial t_s^c}$.

Subcase 3: $t_s^c \in (\frac{1}{f^c}, 1) \implies \lambda_1^c = \lambda_2^c = 0$.

(33) implies $N_s^c \frac{\partial C_s}{\partial N_s^c} = -t_s^c \frac{\partial C_s}{\partial t_s^c}$.

Case 2: $N_s^o > 0, N_s^c = 0 \implies \mu^o = 0$.

(31) implies $p_s^o = \frac{\partial C_s}{\partial N_s^o}$. (32) implies $\frac{\partial C_s}{\partial N_s^c} \geq \frac{p_s^c}{t_s^c f^c}$.

Subcase 1: $t_s^c = 1 \implies \lambda_2^c = 0$

(33) implies $-\lambda_1^c = \frac{\partial C_s}{\partial t_s^c}$. (32) implies $\frac{\partial C_s}{\partial N_s^c} \geq \frac{p_s^c}{f^c}$.

Subcase 2: $t_s^c = \frac{1}{f^c} \implies \lambda_1^c = 0$

(33) implies $\lambda_2^c = \frac{\partial C_s}{\partial t_s^c}$, which contradicts $\frac{\partial C_s}{\partial t_s^c} < 0$.

Subcase 3: $t_s^c \in (\frac{1}{f^c}, 1) \implies \lambda_1^c = \lambda_2^c = 0$.

(33) implies $\frac{\partial C_s}{\partial t_s^c} = 0$, which is a contradiction because $\frac{\partial C_s}{\partial t_s^c} < 0$.

Case 3: $N_s^o = 0, N_s^c > 0 \implies \mu^c = 0$.

(31) implies $\frac{\partial C_s}{\partial N_s^o} \geq p_s^o$. (32) implies $\frac{p_s^c}{t_s^c f^c} = \frac{\partial C_s}{\partial N_s^c}$.

The subcase results of t_s^c are the as Case 1's.

Case 4: $N_s^o = N_s^c = 0$.

Subcase 1: $t_s^c = 1 \implies \lambda_2^c = 0$

(33) implies $-\frac{\partial C_s}{\partial t_s^c} \geq 0$. (31) implies $\frac{\partial C_s}{\partial N_s^o} \geq p_s^o$. (32) implies $\frac{\partial C_s}{\partial N_s^c} \geq \frac{p_s^c}{f^c}$.

Subcase 2: $t_s^c = \frac{1}{f^c} \implies \lambda_1^c = 0$

(33) implies $\lambda_2^c = \frac{\partial C_s}{\partial t_s^c}$, which contradicts $\frac{\partial C_s}{\partial t_s^c} < 0$.

Subcase 3: $t_s^c \in (\frac{1}{f^c}, 1) \implies \lambda_1^c = \lambda_2^c = 0$

(33) implies $\frac{\partial C_s}{\partial t_s^c} = 0$, which contradicts $\frac{\partial C_s}{\partial t_s^c} < 0$.
Q.E.D.

6.2 Proof of Result 2

The Lagrangian is:

$$\mathcal{L} = \sum_j (p_a^j \theta^j - p_s^j) n_a^j - C_a \left(\sum_j \theta^j n_a^j \right) + \mu^o n_a^o + \mu^c n_a^c, \quad (34)$$

where $\mu^c, \mu^o \geq 0$.

The complementary slackness conditions are:

$$\mu^o n_a^o = 0 \quad (35)$$

$$\mu^c n_a^c = 0 \quad (36)$$

FOCs:

$$\forall j : p_a^j \theta^j = p_s^j + \theta^j C'_a \left(\sum_j \theta^j n_a^j \right) - \mu^j \quad (37)$$

$$p_a^c \theta^c = p_s^c + \theta^c C'_a \left(\sum_j \theta^j n_a^j \right) - \mu^c \quad (38)$$

and

$$p_a^o \theta^o = p_s^o + \theta^o C'_a \left(\sum_j \theta^j n_a^j \right) - \mu^o \quad (39)$$

Case 1: $n_a^o > 0, n_a^c > 0$.

$n_a^c > 0 \implies \mu^c = 0$. (38) implies $\theta^c C'_a \left(\sum_j \theta^j n_a^j \right) + p_s^c = \theta^c p_a^c$.

$n_a^o > 0 \implies \mu^o = 0$. (39) implies $\theta^o C'_a \left(\sum_j \theta^j n_a^j \right) + p_s^o = \theta^o p_a^o$.

Case 2: $n_a^o = 0, n_a^c > 0$.

$n_a^c > 0 \implies \mu^c = 0$. (38) implies $\theta^c C'_a \left(\sum_j \theta^j n_a^j \right) + p_s^c = \theta^c p_a^c$.

(39) implies $\theta^o C'_a \left(\sum_j \theta^j n_a^j \right) + p_s^o \geq \theta^o p_a^o$.

Case 3: $n_a^o = 0, n_a^c = 0$.

(38) implies $\theta^c C'_a \left(\sum_j \theta^j n_a^j \right) + p_s^c \geq \theta^c p_a^c$.

(39) implies $\theta^o C'_a \left(\sum_j \theta^j n_a^j \right) + p_s^o \geq \theta^o p_a^o$.

Case 4: $n_a^o > 0, n_a^c = 0$.

$n_a^o > 0 \implies \mu^o = 0$. (39) implies $\theta^o C'_a \left(\sum_j \theta^j n_a^j \right) + p_s^o = \theta^o p_a^o$.

(38) implies $\theta^c C'_a \left(\sum_j \theta^j n_a^j \right) + p_s^c \geq \theta^c p_a^c$.

Q.E.D.

6.3 Proof of Proposition 2

(i) For high-type recipients:

$$\frac{\hat{N}_{\bar{r}}^o}{\hat{t}^o f^o} \leq \frac{\hat{N}_{\bar{r}}^{c,\infty}}{\hat{t}^{c,\infty} f^c} \iff \hat{n}_a^o \leq \frac{R_{\bar{r}}}{R} \hat{n}_a^{c,\infty} \iff \quad (40)$$

$$\frac{\hat{v}_{\bar{r}}^{g,o}}{\hat{\theta}^o} \leq \frac{R_{\bar{r}} \hat{v}_{\bar{r}}^{g,c,\infty}}{R \hat{\theta}^{c,\infty}} \iff \frac{w_{\bar{r}}}{\hat{\theta}^o \hat{p}_a^o} \leq \frac{R_{\bar{r}}}{R} \frac{w_{\bar{r}}}{\hat{\theta}^{c,\infty} \hat{p}_a^{c,\infty}} \quad (41)$$

Since the advertiser's problem gives $\hat{p}_a^o \hat{\theta}^o = \hat{\theta}^o c_a + \hat{p}_s^o$ and $\hat{\theta}^{c,\infty} \hat{p}_a^{c,\infty} = \hat{\theta}^{c,\infty} c_a + \hat{p}_s^{c,\infty}$, and the sender's problem gives $\hat{p}_s^o = c_s^o$ and $\hat{p}_s^{c,\infty} = c_s^c \hat{t}^{c,\infty} f^c$, the last inequality becomes:

$$\hat{\theta}^o c_a + c_s^o \geq \frac{R}{R_{\bar{r}}} (\hat{\theta}^{c,\infty} c_a + c_s^c \hat{t}^{c,\infty} f^c) \iff \hat{\theta}^o \geq \frac{R}{R_{\bar{r}}} \hat{\theta}^{c,\infty} + \frac{R}{R_{\bar{r}}} \frac{c_s^c \hat{t}^{c,\infty} f^c}{c_a} - \frac{c_s^o}{c_a} \quad (42)$$

For low-type recipients, they receive less mail because they don't use the open channel and the mail sent to the censored channel is zero.

(ii) For the total volume received by all recipients:

$$\sum_j \frac{\hat{N}_{\bar{r}}^j + \hat{N}_{\underline{r}}^j}{\hat{t}^j f^j} \leq \frac{\hat{N}_{\bar{r}}^{c,\infty} + \hat{N}_{\underline{r}}^{c,\infty}}{\hat{t}^{c,\infty} f^c} \iff \quad (43)$$

$$\hat{n}_a^o \leq \hat{n}_{\bar{r}}^{c,\infty} + \hat{n}_{\underline{r}}^{c,\infty} (\because \hat{N}_{\bar{r}}^o = \hat{N}_{\underline{r}}^o = 0) \iff \quad (44)$$

$$\frac{\hat{v}_{\bar{r}}^{g,o}}{\hat{\theta}^o} \leq \frac{\hat{v}_{\bar{r}}^{g,c,\infty}}{\hat{\theta}^{c,\infty}} \iff \frac{w_{\bar{r}}}{\hat{\theta}^o \hat{p}_a^o} \leq \frac{w_{\bar{r}}}{\hat{\theta}^{c,\infty} \hat{p}_a^{c,\infty}} \iff \quad (45)$$

$$\hat{\theta}^o c_a + c_s^o \geq \hat{\theta}^{c,\infty} c_a + c_s^c \hat{t}^{c,\infty} f^c \iff \hat{\theta}^o \geq \hat{\theta}^{c,\infty} + \frac{c_s^c \hat{t}^{c,\infty} f^c - c_s^o}{c_a} \quad (46)$$

6.4 Proof of Proposition 3

$$\lim_{f^o \rightarrow \infty} U_r(\hat{v}_r^g, \hat{v}_r^I, \hat{v}_r^{II}) = U_r(\lim_{f^o \rightarrow \infty} \hat{v}_r^g, \lim_{f^o \rightarrow \infty} \hat{v}_r^I, \lim_{f^o \rightarrow \infty} \hat{v}_r^{II}) \quad (47)$$

Since $\hat{U}_r - \lim_{f^o \rightarrow \infty} \hat{U}_r \geq 0$ if (a) $\hat{v}_r^g \geq \lim_{f^o \rightarrow \infty} \hat{v}_r^g$ (with $\hat{v}_r^g > \lim_{f^o \rightarrow \infty} \hat{v}_r^g$), (b) $\hat{v}_r^I \leq \lim_{f^o \rightarrow \infty} \hat{v}_r^I$ (with $\hat{v}_r^I < \lim_{f^o \rightarrow \infty} \hat{v}_r^I$), and (c) $\hat{v}_r^{II} \leq \lim_{f^o \rightarrow \infty} \hat{v}_r^{II}$, we prove each of these inequalities below.

(i) Inequality (a) is $\hat{v}_r^g \geq \lim_{f^o \rightarrow \infty} \hat{v}_r^g$, or:

$$\max_j \left\{ \frac{\kappa_r^j w_r}{p_a^j} \right\} \geq \frac{\hat{\kappa}_r^{c,\infty} w_r}{\hat{p}_a^{c,\infty}} \quad (48)$$

It holds for $w_{\underline{r}} = 0$ because both sides are zero. For $w_{\bar{r}} = 1$, $\hat{\kappa}_{\bar{r}}^o = 1$ by Result 3 and $\kappa_{\bar{r}}^c = \hat{\kappa}_{\bar{r}}^{c,\infty} = 1$ by Assumption 2. (48) becomes:

$$\max_j \left\{ \frac{1}{\hat{p}_a^j} \right\} \geq \frac{1}{\hat{p}_a^{c,\infty}}, \quad (49)$$

which holds with strict inequality since $\hat{p}_a^o < \hat{p}_a^{c,\infty}$. This is because $\hat{p}_a^{c,\infty} = c_a + \frac{\hat{p}_s^{c,\infty}}{\hat{\theta}^{c,\infty}}$ and $\hat{p}_a^o = c_a + \frac{\hat{p}_s^o}{\hat{\theta}^o}$, and $\hat{\theta}^o > \hat{\theta}^{c,\infty}$ ($\frac{R_{\bar{r}}^o}{R^o} = 1$ by Proposition 1) and $\hat{p}_s^o < \hat{p}_s^{c,\infty}$ (where $\hat{p}_s^{c,\infty} = \hat{t}_s^{c,\infty} f^c c_s^{c,\infty}$, $\hat{p}_s^o = c_s^o$, $\hat{t}_s^{c,\infty} f^c \geq 1$ and $c_s^{c,\infty} > c_s^o$).

(ii) Inequality (b) is $\hat{v}_r^I \leq \lim_{f^o \rightarrow \infty} \hat{v}_r^I$, or:

$$\sum_{j \in \{o,c\}} (1 - w_r + w_r \epsilon) \hat{\kappa}_r^j \frac{\hat{N}_r^j}{\hat{t}^j f^j} \leq (1 - w_r + w_r \epsilon) \hat{\kappa}_r^{c,\infty} \frac{\hat{N}_r^{c,\infty}}{\hat{t}^{c,\infty} f^c} \quad (50)$$

For $w_{\underline{r}} = 0$, since $\hat{\kappa}_{\underline{r}}^o = 0$ by Result 3, $\kappa_r^c = \hat{\kappa}_r^{c,\infty} = 1$ by Assumption 2, (50) becomes:

$$\frac{\hat{N}_{\underline{r}}^c}{\hat{t}^c f^c} \leq \frac{\hat{N}_{\underline{r}}^{c,\infty}}{\hat{t}^{c,\infty} f^c},$$

which is true because $\hat{N}_{\underline{r}}^c = 0$ by Proposition 1. If $\hat{N}_{\underline{r}}^{c,\infty} > 0$, the inequality will hold with a strict sign.

For $w_{\bar{r}} = 1$, $\hat{N}_{\bar{r}}^c = 0$ and $\kappa_r^c = \hat{\kappa}_r^{c,\infty} = 1$ implies that (50) becomes (note that if $\epsilon = 0$, (50) will be satisfied because both sides equal zero):

$$\frac{\hat{N}_{\bar{r}}^o}{\hat{t}^o f^o} \leq \frac{\hat{N}_{\bar{r}}^{c,\infty}}{\hat{t}^{c,\infty} f^c} \quad (51)$$

But we already proved in Proposition 2 the necessary and sufficient condition for $\frac{\hat{N}_{\bar{r}}^o}{\hat{t}^o f^o} \leq \frac{\hat{N}_{\bar{r}}^{c,\infty}}{\hat{t}^{c,\infty} f^c}$ to hold.

(iii) Inequality (c) is $\hat{v}_r^{II} \leq \lim_{f^o \rightarrow \infty} \hat{v}_r^{II}$, or:

$$\sum_{j \in \{o,c\}} (1 - \epsilon) w_r \hat{\kappa}_r^j \hat{N}_r^j \left(1 - \frac{1}{\hat{t}^j f^j}\right) \leq (1 - \epsilon) w_r \hat{\kappa}_r^{c,\infty} \hat{N}_r^{c,\infty} \left(1 - \frac{1}{\hat{t}^{c,\infty} f^c}\right) \quad (52)$$

It holds for $w_{\underline{r}} = 0$ because both sides are zero when $\epsilon = 0$. For $w_{\bar{r}} = 1$, since $\hat{\kappa}_{\bar{r}}^o = 1$ by Result 3, $\kappa_r^c = \hat{\kappa}_r^{c,\infty} = 1$ by Assumption 2, (52) becomes:

$$\sum_{j \in \{o,c\}} \hat{N}_r^j \left(1 - \frac{1}{\hat{t}^j f^j}\right) \leq \hat{N}_r^{c,\infty} \left(1 - \frac{1}{\hat{t}^{c,\infty} f^c}\right), \quad (53)$$

which is true because $\hat{N}_r^c = 0$ by Proposition 1 and $\hat{t}^o f^o = 1$.
Q.E.D.

References

- D. Bleichenbacher, E. Gabber, M. Jakobsson, Y. Matias, and A. Mayer. Curbing junk e-mail via secure classification. In *Proc. of the 2nd International Conference on Financial Cryptography*, pages 198–213, 1998.
- Lorrie Faith Cranor and Brian A. LaMacchia. Spam! *Communications of the ACM*, 41(8):74–83, 1998.
- Dave Crocker. Challenges in anti-spam efforts. *The Internet Protocol Journal*, 8(4), 2006.
- Cynthia Dwork and Moni Naor. Pricing via processing or combatting junk mail. In *Advances in Cryptology-CRYPTO 1992*, 1993. Springer-Verlag, 1993, number 740 in Lecture Notes in Computer Science, pp. 139-147.
- Don Evett. Spam statistics 2006, 2006. Available from <http://spam-filter-review.toptenreviews.com/spam-statistics.html>.
- Robert Kraut, Shyam Sunder, Rahul Telang, and James Morris. Pricing electronic mail to solve the problem of spam. *Human-Computer Interaction*, 20(1-2):195–223, 2005.
- Ben Laurie and Richard Clayton. Proof-of-work proves not to work. In *Workshop on Economics and Information Security 2004*, 2004. Available from <http://www.dtc.umn.edu/weis2004/clayton.pdf>.
- Thede Loder, Marshall Van Alstyne, and Rick Wash. An economic response to unsolicited communication. *Advances in Economic Analysis and Policy*, 6(1), 2006. Article 2.
- MessageLabs. Messagelabs intelligence report, 2005.
- MXLogic. Mxlogic reports spam accounts for 67 percent of all email in 2005. Press Release, 22 September 2005. Available from http://www.mxlogic.com/news_events/press_releases/09_22_05_SpamStats.html.

- Geoffrey Parker and Marshall Van Alstyne. Two-sided network effects: A theory of information product design. *Management Science*, 51(10), 2005.
- Juan Carlos Perez. Yahoo and cisco to submit e-mail ID spec to IETF. *NetworkWorld*, 11 July 2005. Available from <http://www.networkworld.com/news/2005/071105-yahoo-cisco.html>.
- J. C. Rochet and J. Tirole. Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003.
- Sophos. *Sophos Security Threat Management Report*. 2005.
- F. W. Taylor. *The Economics of Advertising*. George Allen and Unwin Ltd., 1934.
- Timothy van Zandt. Information overload in a network of targeted communication. *RAND Journal of Economics*, 35(3):542–560, 2004.