Using Uncensored Communication Channels to Divert Spam Traffic*

Benjamin Chiao[†] and Jeffrey MacKie-Mason[‡] 8/31/06
{bchiao, jmm}@umich.edu

Abstract

We offer a microeconomic model of the market for bulk commercial advertising email (the dominant form of spam). We adopt an incentive-centered design approach to develop a simple, feasible improvement to the current email system: an uncensored communication channel. Such a channel could be an email folder or account, to which properly tagged commercial solicitations are routed. We characterize the circumstances under which spammers would voluntarily move much of their spam into the open channel, leaving the traditional email channel dominated by person-to-person, non-spam mail. Our method follows from observing that there is a real demand for unsolicited commercial email, so that 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 show that 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 for all recipients the satisfaction associated with desirable mail received increases, and dissatisfaction associated with undesirable received and desirable mail filtered out decreases.

*We appreciate comments from Nat Bulkley, Nick Economides, Michael Hess, Peter Honeyman, Doug Van Houweling, the members of the Incentive-Centered Design Lab (especially Greg Gamette, Lian Jian, Kil-Sang Kim, John Lin, Anya Osepayshvili, Toinu Reeves, Ben Stearns, and Rick Wash), and participants at the STIET workshop in May 2006. 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

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 modest goal is to identify a particular (but prevalent) subspecies of spam, analyze its ecology, and propose a mechanism that may increase substantially the social welfare by modifying the flows of this type of spam. Our immodest goal is 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 messages sent unsolicited to large numbers of recipients with the goal of inducing a willing, mutually-beneficial purchase by the recipient. 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); and we rule out deceptive email (e.g., "phishing" messages that attempt to trick recipients into revealing valuable personal information such as bank passwords).

Defined as we have done, commercial 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 the fact that we are limiting ourselves to 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 fake Viagra and Rolex knock-offs to pay the spammer's costs.

Casual evidence is consistent with our claim that there is non-trivial demand for much spam: the largest fraction of spam content is commercial advertising for products hard to find through other advertising channels [Cranor and LaMacchia, 1998]. We refer to these as "censored" commercial solicitations, though the censoring is not always explicit or government-supported. Explicitly censored examples include ads for non-prescription providers of regulated drugs, or for providers of knock-off products that intentionally violate copyrights or trademarks of well-known brands. An example that, while not government censored, may have reason to avoid other advertising channels (or may not be accepted by other channels) is (legal) pornography. 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].

Recognizing that *some* recipients want to read spam, while many others evidently do not, we immediately see that one opportunity for

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

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 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 businesspeople, and they will send ads when the net benefit to them exceeds the net cost. What insight do we obtain from this unsurprising observation?[2] 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 should have the same 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.

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.[3] See Figure 1. We then introduce a simple but novel mechanism motivated by the two principles above: an uncensored communication channel through which commercial 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 would accept all appropriately

[2]We 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.

[3]In our current model we focus on the preferences and behaviors of spammers and recipients; we use reduced-form, non-adaptive representations for advertisers and email service providers.

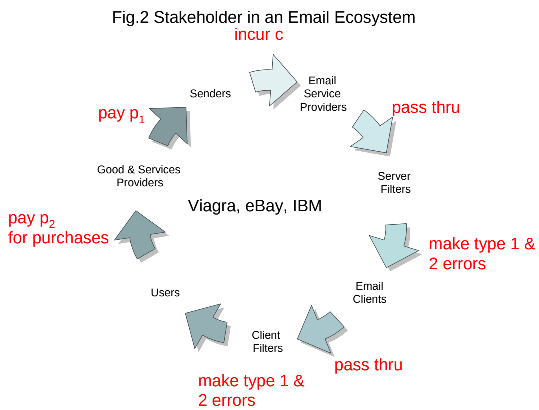

Figure 1: Stakeholders in an e-mail ecosystem.

labeled messages. 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 "spam boxes", and much less at the traditional (censored) channel. See Figure 3.[4]

There should be little dispute that if users could implicitly opt-in for commercial spam by creating an open spam box, the spammers would send mail to that channel. But why would they stop sending to the censored channel? Our conjecture is that if enough of the latent demand for purchasing spam-advertised products is reached through the uncensored spam box channel, then the remaining commercial benefits obtainable from also spamming the traditional censored channel may fall sufficiently low that they no longer justify the incremental costs.

[4]One might argue that the World Wide Web is close to an uncensored channel. If so, why doesn't the Web satisfy the demand for advertising? One obvious reason is that some or many of the products using commercial spam advertising do not want a durable, public presence. If they are moving their web site 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. More generally, we expect there to always be significant demand for "push" advertising in addition to "pull" (search-based) advertising, as 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.).

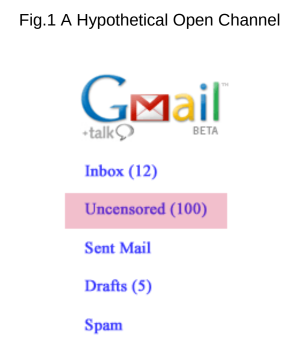

Figure 2: An hypothetical open channel.

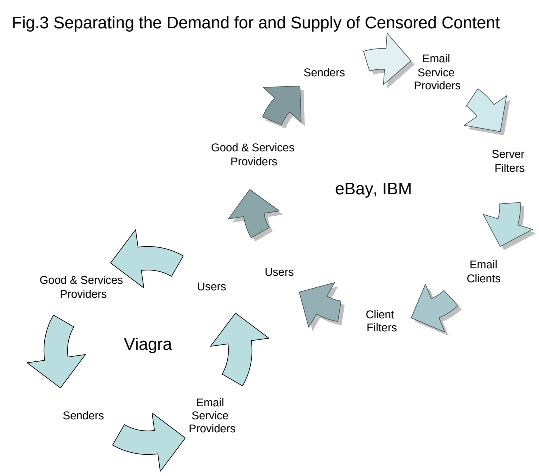

Figure 3: Separating the demand for and supply of unsolicited commercial advertising.

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 purchased 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 advertising: 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 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).

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 [Bleichenbacher et al., 1998], DomainKeys Identified Mail [Perez, 2005], Sender ID or Sender Policy Framework [Crocker, 2006][5], challenge-response [Dwork and Naor, 1993, Laurie and Clayton, 2004], whitelists, and blacklists. See Cranor and LaMacchia [1998] for an overview.

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 falls 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 seem feasible methods to avoid this cycle, given the inevitable and rapid decline in technology costs.

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 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 obtain-

[5]As of now, spam-sending domains are ironically the biggest users of SPF tags [MXLogic, 2005]

ing 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.[6]. Loder et al. [2006] propose an attention-bond mechanism in which a sender deposit a monetary bond to a third-party agent, to be released only if the receiver tells the agent to do so.

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.

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 bulk email service providers to deliver advertisements to email recipients, some of whom in turn willingly choose to purchase the advertised products. 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 unsolicited mail to all recipients, and works only and precisely because recipient preferences are heterogeneous: viz., some recipients want to receive email advertisements.

[6]Two 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 emails 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.

3.1 The Recipients' Problem

Recipient t is either of type (h)igh or (l)ow. The type is the intensity of preference for censored content. Channel j is either (o)pen or (c)ensored.[7]

Senders are defined by two attributes: whether they send mass or targeted mail, and whether they send censored or uncensored content.[8]

The first attribute is mainly a cost attribute. Specifically, mass-mail satisfies the following characteristics of distribution and information costs:

- Low distribution cost. The population of senders could in principle be distributed continuously according to how efficient they are in sending mass-mail. For simplicity and to a great extent a description of the real world, we assume however that either a sender can or cannot send mass-mail. The ability to send mass-mail refers to the ability to sending tens of thousands of mail per day.

- Revenue invariance to randomization. A mass-mail does not imply that copies are necessarily identical. In fact, a common strategy is for a mass-mail sender to randomize some uninformative text in an otherwise identical message to fool content filters. It is equivalent, for our purpose, if a sender randomizes over information content (e.g., sends some medication messages and some mortgage loan messages), but obtain the same expected revenue from each.

By both of the above characteristics, mass mail always appear in multiple and sometimes almost identical copies in a given recipient's inbox. This higher substitutability implies that there is *a low type II error cost* associated with mass mail. That is, mass mail wrongly filtered will cause much less inconvenience than the counterpart of targeted-mail, even for those recipients who want mass-mail. The converse is not true. Some recipients prefer even to neglect targeted mail from some people they know. We are going to assume, however, that on average there is a much lower type II error cost associated with mass mail.

We do not require that it is possible to identify that mail is sent by a bulk service provider. It is easy to fool general purpose filters of the

[7] We call the currently available standard email channel *censored* because service provider filters and domain-blocking rules are ubiquitous.

[8] By *censored* content we mean content of a type that conventional email service providers routinely attempt to filter out of the recipient's email stream. Such content may or may not be illegal, and the filtering efforts generally will be imperfect. Thus, as we make explicit below, some *censored* content may be unfiltered, and thus be received.

identity of the sender, and the recipient often won't know until after incurring the cost of viewing the message. In practice, much spam can be automatically identified as being sent from a bulk provider, but our results are robust as long as considerable spam cannot.

The second sender attribute is whether it sends content of a type that is censored (if recognized) by the email service provider (ESP).[9] Content-based filtering can rely on any available information headers and body text. For example, Gmail, Hotmail and Yahoo! usually filter adult content and all mail from some blacklisted senders' (usually based on IP addresses). On the other hand, we assume that senders can, at a cost, disguise content to some degree.

In all, we identify four types of mail:

Censored-content mass Examples include Viagra and erotic content advertisements.

Censored-content targeted Examples include personalized adult materials, perhaps sent by a pay subscription service.

Uncensored-content mass Examples include advertisements from conventional booksellers, non-profit fundraisers, and other legal and less socially objectionable purveyors.

Uncensored-content targeted Examples include personal correspondence.

Our design goal was to develop a social welfare-increasing mechanism that induces censored-content mass mailers to reduce the supply of their messages delivered to the currently standard email channel (the *censored* channel). Therefore, we simplify by assuming that mass mailers send only censored content, and targeted mailers send only uncensored content.[10]

To model the user problem we suppose that recipient r chooses which channel(s) to read in order to maximize utility, which depends on the quantity of various categories of email:

$$U^r(\text{desired mail received, undesired mail received, desired mail not received}) \tag{1}$$

[9]Recipient censorship (with, for example, personal spam filters) are not very important to our central results, as long as the value of spam that evades these filters is, on average, negative to a segment of the population.

[10]There are interesting research questions associated with the other two email types as well, but they fall outside the scope of our present analysis. Adding them to our model for the questions we ask in this paper would complicate notation and proofs, but would not change the qualitative results.

The utility function is increasing in the first argument, and decreasing in the others. Before explaining the arguments above, we introduce further notation.

Assume that for all recipients, there is a (perhaps small) fraction ϵ of uncensored mail that is not desired. We assume that individuals either desire (all) censored-content mail in a given channel or not, and use the indicator ϕ_t^j to represent those preferences. If a recipient of type $t \in \{h, l\}$ desires censored-content mail in channel j, then $\phi_t^j = 1$; otherwise $\phi_t^j = 0$.[11] We assume that only high type recipients put a positive value on censored content ($\phi_l^j = 0, \phi_h^j = 1$). 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".

Then the first argument of the full utility function (1), desired mail received, becomes:

$$(1-\epsilon)*\text{unfiltered uncensored mail}+\text{unfiltered censored mail} \times \phi_t^j \quad (2)$$

The second argument of utility function (1), undesired mail received, becomes:

$$\epsilon*\text{unfiltered uncensored mail}+\text{unfiltered censored mail} \times (1-\phi_t^j) \quad (3)$$

The third argument of the utility function (1), desired mail not received, becomes:

$$(1-\epsilon)*\text{filtered uncensored mail}+\text{filtered censored mail} \times \phi_t^j \quad (4)$$

In the censored channel filtering technology is designed to distinguish between censored and uncensored content, but it does so imperfectly. Each sender knows that the filter has a strength of $\gamma^c \in [1, \infty)$ for censored content, and strength $\hat{\gamma}^c \in [1, \infty)$ for uncensored content, with $\gamma^c > \hat{\gamma}^c$. 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, $\gamma^o = 1$.

Sender s can make an effort to disguise its content to reduce the filter's success rate. We let sender s choose a disguise level, $d_s^j \in [\frac{1}{\gamma^j}, 1]$,

[11] We have an asymmetry between the fraction of desirable censored- and uncensored-content mail in a channel: recipients may not want 100% of the uncensored mail sent to them in a channel, but if they want any censored-content mail, then want all of it. We do this to simplify the algebra, without losing anything qualitatively important. In both cases, not all mail is desired: for uncensored, each individual may not want some; for censored, some individuals don't want any. Thus, there is the possibility of both Type I and Type II errors for each.

for mail sent to channel j, where d_s^j is a multiplicative factor adjusting filter strength. If $d_s^j = 1$, disguising has no impact and the effective filter strength is the technological strength γ^j. If $d_s^j = 1/\gamma^j$, the effective filter strength is one, which is to say, all content passes through unfiltered. Disguising is costly, so we let $d_s^o = 1$ (no effort to disguise in the open channel).

We let n denote the volume of censored-content mail, and $\hat{n}$ denote the volume of uncensored-content mail. Then, for some given censored-content mail volume sent to recipient r in the censored channel, n_r^c, the portion that actually reaches the recipient is $\frac{n_r^c}{d^c\gamma^c}$, where d^j is the disguise level associated with n_r^j. Note that we assume that there is no need to disguise uncensored content.

In our informal specification (1), recipient utility depends on the undifferentiated volume of various mail categories. However, by introducing content disguising, we cannot avoid another dimension of quality: the value of a given type of mail to a recipient will now also depend on how informative it is, which generally will be inversely proportional to the amount of disguising the sender does. That is, cluttering a message with extraneous garbage text to get past a filter also makes it difficult for the recipient to find the useful information. Therefore, we allow utility to depend on the informativeness-adjusted volume of email received. To adjust for message informativeness after disguising, we introduce an adjustment function b, which is increasing in the effort made to disguise censored-content mail.

We define $\kappa_r^j = 1$ if recipient r uses channel j, zero otherwise.

Now we can formally express the utility function (1). The first argument, which is informativeness-adjusted desired mail received, becomes:

$$u^r_{\text{desired received}} = \sum_{j\in\{o,c\}} \underbrace{(1-\epsilon)\kappa_r^j \frac{\hat{n}_r^j}{\hat{\gamma}^j}}_{\text{uncensored content mail}} + \sum_{j\in\{o,c\}} \underbrace{\phi_t^j \kappa_r^j \frac{b(d^j) n_r^j}{d^j \gamma^j}}_{\text{censored content mail}} \tag{5}$$

in which the first term is (desirable) unfiltered uncensored-content mail, and the second term is unfiltered, censored-content, and disguised mail for high type recipients (i.e., those who find it desirable).

The second argument of the utility function (1), which is informativeness-adjusted undesired mail received, becomes:

$$u^r_{\text{Type I errors}} = \sum_{j\in\{o,c\}} \epsilon \kappa_r^j \frac{\hat{n}_r^j}{\hat{\gamma}^j} + \sum_{j\in\{o,c\}} (1-\phi_t^j)\kappa_r^j \frac{b(d^j) n_r^j}{d^j \gamma^j} \tag{6}$$

in which the first term is undesirable unfiltered uncensored mail, and the second term is unfiltered, censored-content, disguised mail for low type recipients (who give it a negative value).

The third argument of utility function (1), desired mail not received, becomes:

$$u^r_{\text{Type II errors}} = \sum_{j \in \{o,c\}} (1-\epsilon)\kappa^j_r \hat{n}^j_r (1 - \frac{1}{\hat{\gamma}^j}) + \sum_{j \in \{o,c\}} \phi^j_t \kappa^j_r b(d^j) n^j_r (1 - \frac{1}{d^j \gamma^j}) \tag{7}$$

where the first term is desired filtered uncensored mail, and the second term is filtered censored-content mail for high type recipients.[12]

As a special case, we assume that the censored channel is essential so that every recipient uses it: $\kappa^c_r = 1$. Then, given the filter strengths, disguise levels, email volume and actions of other recipients, recipient r makes a binary choice whether to read mail in the open channel, $\kappa^o_r \in \{0, 1\}$, by maximizing:

$$U^r(u^r_{\text{desired received}}, u^r_{\text{Type 1 errors}}, u^r_{\text{Type 2 errors}}) \tag{8}$$

Proposition 1 *If every recipient uses the censored channel and there is no uncensored-content mail in the open channel* ($\hat{n}^o_r = 0$)*, then recipients who have a positive value for censored contents, and only they, will use also the open channel.*

Proof. The result is obtained straightforwardly from the three components of (dis)utility, (5)–(6). First, for a recipient who finds censored content undesirable ($\phi^j_l = 0$), reading the open channel provides no benefit, but creates disutility by increasing the amount of objectionable mail (see the second summand in (6)). For a recipient who values censored content, reading mail in the open channel increases the second summand in (5) (desired mail received). It has no effect on Type I errors (6). Likewise it has no effect on Type II errors (7) because for the open channel $d^j = \gamma^j = 1$, so the second summand is zero when $j = o$. ■

Thus, if an open channel is introduced, h-type recipients will use it to obtain benefit from desired commercial spam, but l-types, who do not want spam, will not (as long as personal senders do not start sending (much) to the open channel). We now turn to senders to find the equilibrium behavior of spammers when an open channel is introduced, after which we analyze the welfare effects of an open channel.

[12] We could elaborate by allowing Type II errors associated with targeted mail to be more annoying.

3.2 The Sender's Problem

We will describe in detail the cost and revenue functions of the censored-content mass-mail senders only. This is because the focus of the paper is to move the supply of and demand for censored-content mass-mail out of the current email system.

The total cost function for mass-email sender $m, c_m(n_m^o, n_m^c, d_m^c)$, reflects the costs of generating the email volumes, and of disguising mail sent to the censored channel. The disguise cost is captured by $\partial c_m/\partial d_m^j < 0$, and the volume generating cost by $\partial c_m/\partial n_m^j > 0$.[13] We allow for economies of scale in the sense of sub-additivity, $c_m(n_m^o, 0, d_m^c) + c_m(0, n_m^o, d_m^c) > c_m(n_m^o, n_m^c, d_m^c)$, and cost complementarity (ie., $\frac{\partial^2 c_m}{\partial n_m^j \partial n_m^i} < 0, i \neq j$). To be concrete, we specify $c_m(n_m^o, n_m^c, d_m^c) = FC_m + g_m(d_m^c) + \delta n_m^o n_m^c + \frac{1}{2}(n_m^o)^2 + \frac{1}{2}(n_m^c)^2$, in which $g_m(d_m^c) = \frac{1}{d_m^c} - 1$[14], so that the cost of no disguising ($d_m^c = 1$) is $g_m(1) = 0$. Cost complementarity and subadditivity are both ensured by letting $\delta < 0$.[15] We also assume a regularity condition of $\delta^2 < 1$.

On the revenue side, senders are price takers. Sellers of censored goods or other legitimate goods pay them for solicitations. Let p^j be the advertising charge per disguised email ($\frac{n_m^j}{d_m^j \gamma^j}$) reaching the users in channel j[16].

[13] Rather than having a zero marginal cost as commonly asserted, spammers incur cost to renew technologies, which depreciate quickly, to generate spam. For example, zombies (ie. 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).

[14] We could have used a decreasing marginal cost function such as $g(d_m^c) = \frac{1}{(d_m^c)^2} - 1$.

[15] Cost complementarity follows from $\delta < 0$ because $\frac{\partial c_m}{\partial n_m^c} = \delta n_m^o + n_m^c$, and $\frac{\partial c_m}{\partial n_m^o} = \delta n_m^c + n_m^o$. Subadditivty does as well because

$$\begin{aligned}
& c_m(n_m^o, n_m^c, d_m^c) - [c_m(n_m^o, 0, d_m^c) + c_m(0, n_m^c, d_m^c)] \\
=& FC_m + g_m(d_m^c) + \delta n_m^o n_m^c + \frac{1}{2}(n_m^o)^2 + \frac{1}{2}(n_m^c)^2 - [FC_m + \\
& \frac{1}{2}(n_m^o)^2 + FC_m + g_m(d_m^c) + \frac{1}{2}(n_m^c)^2] \\
=& \delta n_m^o n_m^c - FC_m < 0.
\end{aligned}$$

[16] In practice, there is a volume discount (that might or might not 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 (a proxy of $\frac{n_m^j}{d_m^j \gamma^j}$) 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.

On a practical level, the sender chooses whether to send to the censored or the open channel (or both). If 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 closed channel, the sender does not tag the message, and in fact may expend some effort to disguise the content. We assume that mail send is distributed uniformly to the recipients in a given channel.

Given the prices and filter strengths, sender m chooses (n^o_m, n^c_m, d^c_m) to maximize:

$$\pi_m(n^o_m, n^c_m, d^c_m) = p^o n^o_m + \frac{p^c n^c_m}{d^c_m \gamma^c} - c_m(n^o_m, n^c_m, d^c_m) \tag{9}$$

s.t.

$$d^c_m \in [\frac{1}{\gamma^c}, 1]. \tag{10}$$

Proposition 2 *Consider three cases. Case (a):* $p^o \leq \frac{p^c}{\delta\gamma^c} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$*; case (b):* $p^o \geq \frac{p^c}{\delta} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$*; case (c): Otherwise. If* $p^o = p^c = 0$*,*

$$d^{*c}_m = 1; n^{*o}_m = 0; n^{*c}_m = 0 \tag{11}$$

Else, the best responses of sender m are:
Case (a):

$$d^{*c}_m = 1; n^{*o}_m = \frac{1}{1-\delta^2}(p^o - \frac{\delta p^c}{\gamma^c}); n^{*c}_m = \frac{1}{1-\delta^2}(\frac{p^c}{\gamma^c} - \delta p^o) \tag{12}$$

Case (b):

$$d^{*c}_m = \frac{1}{\gamma^c}; n^{*o}_m = \frac{1}{1-\delta^2}(p^o - \delta p^c); n^{*c}_m = \frac{1}{1-\delta^2}(p^c - \delta p^o) \tag{13}$$

Case (c):

$$d^{*c}_m = \frac{(p^c)^2}{(1-\delta^2)(\gamma^c)^2 + \delta p^o p^c \gamma^c}; \tag{14}$$

$$n^{*o}_m = p^o - \delta\frac{\gamma^o}{p^c}; n^{*c}_m = \frac{\gamma^c}{p^c} \tag{15}$$

Proof. See Appendix 1. ■

Corollary 1 *As long as either p^o or p^c is (or both are) strictly positive, a non-zero quantity of mass email will be sent to both channels.*

Proposition 3 *When either p^o or p^c is (or both are) strictly positive, the optimal solutions n_m^{*o}, n_m^{*c} and d_m^{*c} exhibit the following characteristics:*
(1) d_m^{*c} is i) decreasing in γ^c in cases b and c, independent otherwise; ii) increasing in p^o and p^c in case c, independent otherwise; iii) decreasing in δ in case c, independent otherwise.
(2) n_m^{*o} and n_m^{*c} are i) increasing in both p^o and p^c in cases a and b; ii) decreasing in δ in cases a and b; iii) decreasing in γ^c in case a, and independent in case b. For case (c), n_m^{*o} is increasing in p^o and decreasing in p^c. For case (c), n_m^{*c} is decreasing in p^c and is independent of p^o.

Proof. By differentiation of each of the best responses in Proposition 2. ■

3.3 Welfare

Proposition 4 *The welfare of censored-content mass-mail senders and all recipients will be unchanged or increased when there is an open channel if the following assumptions hold: (1) there is no censored-content targeted-mail; (2) there is no uncensored-content mail in the open channel; (3) $b(d^j) \equiv (d^j)^2$; (4) the marginal revenue of sending uncensored-content mail to the censored channel does not change with or without the open channel; (5) $p^o \geq p^c$.*(6) *Either p^o or p^c equal to zero.*

Proof. See Appendix 6.2 ■

4 Implementation Issues

We emphasize that our proposal is a starting point. There are implementation issues, which are outside the scope of this research, that must be addressed:

- Will the total trade volume of censored goods increase? Is it reasonable to assume that the open channel simply shifts the supply of such goods from other outlets?

- What is the magnitude of the marginal exposure of pornography for minors in the open channel? Have they already been exposed significantly by websites on the Internet? Should we add minimal censorship to the open channel by blocking sexually explicit images or requiring credit card numbers to access the open channel? Or should we block at least some contents with viruses and worms? Will the main argument still hold as long as the open channel is significantly less censored than other channels? More generally, what are the social implications if it is easier to obtain counterfeit products or pirated software because of the open channel?

- The open channel is a typical problem of two-sided markets, which need both the sides of senders and recipients. Is it desirable for the large email service providers to unilaterally opt-in for all the recipients (so at least one side of the market is on board)?[17] Currently, Gmail lists side-by-side some advertisements even for spam messages. Will the possibly increased email volume (at least email with censored contents) be sufficient incentives for the private provision of the open channel? How many providers' adoptions do we need for the open channel to be effective? Are the customers willing to switch to the few adopters?

5 Conclusions

We propose a principled approach to developing and analyzing spam policies. Our approach is grounded in an economic, rational choice characterization of the choices made by spammers and recipients. Our novel insight is to induce the suppliers for and demanders of commercial spam to move out of the current email system (a censored channel), by providing an open channel in which those who want the advertisements can find them. As a corollary benefit, resources are not wasted on unproductive content disguising, and readers receive higher quality (more informative) ads.

Technical filters and legal rules raise the cost of delivering spam to readers. Costs are borne by advertisers (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. Methods that channel communications more directly to those who want them would lower costs on both sides and be welfare improving.

In our mathematical model, we have shown that all email recipients are better off with the introduction of such open channel: only recipients wanting spam will use the open channel enjoying the less disguised messages, and for all recipients the satisfaction associated with desirable mail received increases, and dissatisfaction associated with undesirable 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 the other well-known tools (technical, legal and economic), may contribute to a reduction in the flow of low-information, unsolicited bulk email. The ultimate solution, simple economics predicts, is for the value of purchasing stimulated by spam to fall sufficiently low that

[17] Gmail, Yahoo!, and Hotmail are three largest online email service providers, each with a market share close to 1/3.

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 (tempt them with the expectation of higher quality, more informative ads to help them find the products they want), the purchasing value remaining in the traditional, filtered channel may drop sufficiently to start discouraging 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 an efficient free market — the open channel — for those who do not want to dispose of spam. But 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 don't 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 readers move to the open channel.

Of course, not all spam is designed to deliver informative advertising messages to willing customers. A significant portion of spam is intended to deceive readers (e.g., phishing and other scams), and other spam messages are intended to *persuade* readers who may not have previously thought they wanted to purchase a spam-advertised product (and thus, who would not read the messages in the uncensored advertising channel). We do not suggest that our proposal will have a direct effect on the quantity of misleading spam email (it might affect persuasive advertising because a large fraction of those susceptible to this may already be inclined to read the uncensored and more informative advertising channel).

An open advertising channel is possible at low cost, and it seems evident would make email users at least weakly better off (no worse off) than the status quo. 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 Proposition 2

The sender's profit function is

$$\pi_m(n_m^o, n_m^c, d_m^c) = p^o n_m^o + \frac{p^c n_m^c}{d_m^c \gamma^c} - c_m(n_m^o, n_m^c, d_m^c), \tag{16}$$

so, the Lagrangian is:

$$\mathcal{L} = \pi(\cdot) - \lambda_1^c(d_m^c - 1) + \lambda_2^c(d_m^c - \frac{1}{\gamma^c}) + \mu^o n_m^o + \mu^c n_m^c \tag{17}$$

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

The complementary slackness conditions are:

$$\lambda_1^c(d_m^c - 1) = 0 \tag{18}$$

$$\lambda_2^c(d_m^c - \frac{1}{\gamma^c}) = 0 \tag{19}$$

$$\mu^o n_m^o = 0 \tag{20}$$

$$\mu^c n_m^c = 0 \tag{21}$$

FOCs:

$$p^o = \frac{\partial c_m}{\partial n_m^o} - \mu^o = \delta n_m^c + n_m^o - \mu^o$$

$$\implies n_m^o = p^o - \delta n_m^c + \mu^o \tag{22}$$

$$\frac{p^c}{d_m^c \gamma^c} = \frac{\partial c_m}{\partial n_m^c} - \mu^c = \delta n_m^o + n_m^c - \mu^c$$

$$\implies n_m^c = \frac{p^c}{d_m^c \gamma^c} - \delta n_m^o + \mu^c \tag{23}$$

$$\frac{-p^c n_m^c}{(d_m^c)^2 \gamma^c} - \lambda_1^c + \lambda_2^c = \frac{\partial c_m}{\partial d_m^c} = g_m'(d_m^c) = -(d_m^c)^{-2}$$

$$1 = \frac{p^c n_m^c}{\gamma^c} + (\lambda_1^c - \lambda_2^c)(d_m^c)^2 \tag{24}$$

$$\implies d_m^c = \left(\frac{1 - \frac{p^c n_m^c}{\gamma^c}}{\lambda_1^c - \lambda_2^c}\right)^{1/2}, \lambda_1^c \neq \lambda_2^c \tag{25}$$

Combining (22) and (23):

$$\begin{aligned} n_m^o &= p^o - \delta(\frac{p^c}{d_m^c \gamma^c} - \delta n_m^o + \mu^c) + \mu^o \\ &= \frac{1}{1-\delta^2}[p^o - \delta(\frac{p^c}{d_m^c \gamma^c} + \mu^c) + \mu^o] \end{aligned} \tag{26}$$

$$\begin{aligned} n_m^c &= \frac{p^c}{d_m^c \gamma^c} - \delta(p^o - \delta n_m^c + \mu^o) + \mu^c \\ &= \frac{1}{1-\delta^2}[\frac{p^c}{d_m^c \gamma^c} - \delta(p^o + \mu^o) + \mu^c] \end{aligned} \tag{27}$$

Before doing more substitutions in the above nonlinear equations to solve for d_m^c, n_m^o, n_m^c more explicitly, we first see if we could eliminate some cases below.

Case 1: $n_m^{*o}, n_m^{*c} > 0 \implies \mu^o = \mu^c = 0$.
From (26),

$$n_m^o = \frac{1}{1-\delta^2}[p^o - \delta(\frac{p^c}{d_m^c \gamma_m^c})] \tag{28}$$

From (27),

$$n_m^c = \frac{1}{1-\delta^2}[\frac{p^c}{d_m^c \gamma^c} - \delta p^o] \tag{29}$$

<u>Subcase 1:</u> $d_m^{*c} = 1 \implies \lambda_2^c = 0$
From (28),

$$n_m^{*o} = \frac{1}{1-\delta^2}[p^o - \delta(\frac{p^c}{\gamma_m^c})] \tag{30}$$

From (29),

$$n_m^{*c} = \frac{1}{1-\delta^2}[\frac{p^c}{\gamma^c} - \delta p^o] \tag{31}$$

From (24),

$$\lambda_1^c = \frac{\gamma^c - p^c n_m^{*c}}{\gamma^c} \tag{32}$$

From (32), $\lambda_1^c > 0 \iff$

$$\gamma^c > p^c n_m^{*c} \tag{33}$$

$$\gamma^c > \frac{p^c}{1-\delta^2}(\frac{p^c}{\gamma^c} - \delta p^o) \tag{34}$$

From (32), $\lambda_1^c = 0 \iff$

$$\gamma^c = p^c n_m^{*c}$$

$$\gamma^c = \frac{p^c}{1-\delta^2}(\frac{p^c}{\gamma^c} - \delta p^o)$$

Therefore, subcase 1 is admissible when $\gamma^c \geq \frac{p^c}{1-\delta^2}(\frac{p^c}{\gamma^c} - \delta p^o)$, or equivalently $p^o \leq \frac{p^c}{\delta\gamma^c} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$.[18]

Subcase 2: $d_m^{*c} = \frac{1}{\gamma^c} \implies \lambda_1^c = 0$

From (28),

$$n_m^{*o} = \frac{1}{1-\delta^2}[p^o - \delta p^c] \tag{35}$$

From (29),

$$n_m^{*c} = \frac{1}{1-\delta^2}[p^c - \delta p^o] \tag{36}$$

From (24),

$$1 = \frac{p^c n_m^{*c}}{\gamma^c} - \frac{\lambda_2^c}{(\gamma^c)^2} \tag{37}$$

$$\lambda_2^c = \gamma^c p^c n_m^{*c} - (\gamma^c)^2 \tag{38}$$

From (38), $\lambda_2^c > 0 \iff$

$$\gamma^c p^c n_m^{*c} - (\gamma^c)^2 > 0 \tag{39}$$

$$p^c n_m^{*c} > \gamma^c \tag{40}$$

$$\frac{p^c(p^c - \delta p^o)}{1-\delta^2} > \gamma^c \tag{41}$$

From (38), $\lambda_2^c = 0 \iff$

$$p^c n_m^{*c} = \gamma^c \tag{42}$$

$$\frac{p^c(p^c - \delta p^o)}{1-\delta^2} = \gamma^c \tag{43}$$

Therefore, subcase 2 is admissible when $\gamma^c \leq \frac{p^c}{1-\delta^2}(p^c - \delta p^o)$ or equivalently $p^o \geq \frac{p^c}{\delta} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$.

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

Equation (24) and the premises for this subcase imply that:

$$\gamma^c = n_m^c p^c \tag{44}$$

[18] The inequality reverses direction because we multiplied both sides by $\delta < 0$.

Substitute (44) into (29) to get[19]:

$$d_m^{*c} = \frac{(p^c)^2}{(1-\delta^2)(\gamma^c)^2 + \delta p^o p^c \gamma^c} \tag{45}$$

The solution for n_m^{*c} is already available from (44) (or can be equivalently obtained by substituting (45) into (29)):

$$n_m^{*c} = \frac{\gamma^c}{p^c} \tag{46}$$

Substitute (46) into (22) to get

$$n_m^{*o} = p^o - \delta\frac{\gamma^c)}{p^c} \tag{47}$$

Case 2: $n_m^{*o} > 0, n_m^{*c} = 0$
From (23),

$$\begin{aligned} n_m^c &= \frac{p^c}{d_m^c \gamma^c} - \delta n_m^o + \mu^c \\ 0 &= \frac{p^c}{d_m^c \gamma^c} - \delta n_m^o + \mu^c \end{aligned}$$

Since the right hand side is non-negative, the only permissible values are $p^c = \mu^c = n_m^{*o} = 0$,which contradicts with $n_m^{*o} > 0$.
Case 3: $n_m^{*o} = 0, n_m^{*c} > 0$
From (22),

$$n_m^o = p^o - \delta n_m^c + \mu^o \tag{48}$$
$$0 = p^o - \delta n_m^c + \mu^o \tag{49}$$

Since the right hand side is non-negative, the only permissible values are $p^o = \mu^o = n_m^{*c} = 0$,which contradicts with $n_m^{*c} > 0$.

19

$$\begin{aligned} n_m^c &= \frac{1}{1-\delta^2}\left(\frac{p^c}{d_m^c \gamma^c} - \delta p^o\right) \\ \frac{\gamma^c}{p^c} &= \frac{1}{1-\delta^2}\left(\frac{p^c}{d_m^c \gamma^c} - \delta p^o\right) \\ \delta p^o + (1-\delta^2)\frac{\gamma^c}{p^c} &= \frac{p^c}{d_m^c \gamma^c} \\ d_m^{*c} &= \frac{(p^c)^2}{(1-\delta^2)(\gamma^c)^2 + \delta p^o p^c \gamma^c} \end{aligned}$$

Case 4: $n_m^{*o} = n_m^{*c} = 0$.
From (22),

$$n_m^o = p^o - \delta n_m^c + \mu^o \tag{50}$$
$$0 = p^o + \mu^o \tag{51}$$

Since the right hand side is non-negative, the only permissible values are $p^o = \mu^o = 0$.
From (23),

$$n_m^c = \frac{p^c}{d_m^c \gamma^c} - \delta n_m^o + \mu^c \tag{52}$$
$$0 = \frac{p^c}{d_m^c \gamma^c} + \mu^c \tag{53}$$

Since the right hand side is non-negative, the only permissible values are $p^c = \mu^c = 0$.
$n_m^{*c} = 0$ implies that (25) gives:

$$d_m^{*c} = (\frac{1}{\lambda_1^c - \lambda_2^c})^{1/2} \tag{54}$$

Subcase 1: If $\lambda_1^c = 0$ and $\lambda_2^c = 0$, (24) implies a contradiction because:

$$\frac{p^c n_m^{*c}}{\gamma^c} = 1 \tag{55}$$
$$n_m^{*c} = \frac{\gamma^c}{p^c} \neq 0 \tag{56}$$

Subcase 2: If $\lambda_1^c = 0$ and $\lambda_2^c > 0$, it gives a contradiction of d_m^{*c} being negative by (54).
Subcase 3: If $\lambda_1^c > 0 \implies d_m^c = 1$ (see (18)).
Therefore, case 4 is admissible when

$$d_m^c = 1; p^o = p^c = 0 \tag{57}$$

6.2 Proof of the Welfare Gain of the Recipients

Assume the interchangeability of the limit signs for U^r :

$$\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} U^r(u^r_{\text{desired received}}, u^r_{\text{Type 1 errors}}, u^r_{\text{Type 2 errors}}) \tag{58}$$
$$= U^r(\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{desired received}}, \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 1 errors}}, \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 2 errors}}) \tag{59}$$

where (use ∞ as a superscript to denote the best responses of the senders when $\gamma^o, \hat{\gamma}^o \longrightarrow \infty$):

$$\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{desired received}} = (1-\epsilon)\frac{\kappa^c_r \hat{n}^{c,\infty}_r}{\hat{\gamma}^c} + \frac{\phi^c_t \kappa^c_r b(d^{c,\infty}) n^{c,\infty}_r}{d^{c,\infty}\gamma^c} \tag{60}$$

$$\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 1 errors}} = \epsilon\frac{\kappa^c_r \hat{n}^{c,\infty}_r}{\hat{\gamma}^c} + \frac{(1-\phi^c_t)\kappa^c_r b(d^{c,\infty}) n^{c,\infty}_r}{d^{c,\infty}\gamma^c} \tag{61}$$

$$\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 2 errors}} \tag{62}$$

$$= (1-\epsilon)\kappa^c_r \hat{n}^{c,\infty}_r (1-\frac{1}{\hat{\gamma}^c}) + \phi^c_t \kappa^c_r b(d^{c,\infty}) n^{c,\infty}_r (1-\frac{1}{d^{c,\infty}\gamma^c}) \tag{63}$$

Since $U^r - \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} U^r \geq 0$ if (a) $u^r_{\text{desired received}} \geq \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{desired received}}$,(b) $u^r_{\text{Type 1 errors}} \leq \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 1 errors}}$, and (c) $u^r_{\text{Type 2 errors}} \leq \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 2 errors}}$, we are going to prove each of these inequalitities in three parts.

(i) Inequality (a) is $u^r_{\text{desired received}} \geq \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{desired received}}$, or equivalently:

$$\sum_{j\in\{o,c\}} (1-\epsilon)\kappa^j_r \frac{\hat{n}^j_r}{\hat{\gamma}^j} + \sum_{j\in\{o,c\}} \phi^j_t \kappa^j_r \frac{b(d^j) n^j_r}{d^j \gamma^j}$$
$$\geq (1-\epsilon)\frac{\kappa^c_r \hat{n}^{c,\infty}_r}{\hat{\gamma}^c} + \frac{\phi^c_t \kappa^c_r b(d^{c,\infty}) n^{c,\infty}_r}{d^{c,\infty}\gamma^c} \tag{64}$$

As in Proposition 1, if we assume that there is no uncensored-content mail in the open channel (implying $(1-\epsilon)\kappa^o_r \frac{\hat{n}^o_r}{\hat{\gamma}^o} = 0$), then it implies that the senders of such mail already find it profit maximizing by sending only to the censored channel. Since the marginal cost function has not changed, as long as the marginal revenue of sending uncensored-content mail to the censored channel does not change with or without the open channel, such senders will not change the values of decision variables chosen even when the open channel disappears because the first order condition of equations of equating marginal revenue and marginal cost do not change. This implies that $(1-\epsilon)\kappa^c_r \frac{\hat{n}^c_r}{\hat{\gamma}^c} = (1-\epsilon)\frac{\kappa^c_r \hat{n}^{c,\infty}_r}{\hat{\gamma}^c}$. Then it is equivalent to proving that:

$$\frac{\phi^c_t \kappa^o_r n^o_r}{\gamma^o} \geq \frac{\phi^c_t \kappa^c_r b(d^{c,\infty}) n^{c,\infty}_r}{d^{c,\infty}\gamma^c} \tag{65}$$

The above is true for $\phi_t^o = 0$ because both sides are zero. For $\phi_t^o = 1$, because $\kappa_r^o = \kappa_r^c = 1$ by Proposition 1 and $\gamma_s^o = 1$, it is sufficient to prove that $n_r^o \geq n_r^{c,\infty}$ where r is a high type recipient because:

$$n_r^o \geq \frac{b(d^{c,\infty})n_r^{c,\infty}}{d^{c,\infty}\gamma^c} \iff \tag{66}$$

$$n_r^o \geq n_r^{c,\infty} \text{ because } d^{c,\infty}\gamma^c \geq 1 \text{ and } b(d^{c,\infty}) \leq 1 \iff \tag{67}$$

To prove that $n_r^o \geq n_r^{c,\infty}$. Let's first look at the censored-content mass-mail sender's problem. For n_m^o, the marginal revenue is p^o and the marginal cost is $\frac{\partial c_m}{\partial n_m^o} = \delta n_m^c + n_m^o$. For $n_m^{c,\infty}$, the marginal revenue is $\frac{p^c}{d_m^{c,\infty}\gamma^c}$ and the marginal cost is $n_m^{c,\infty}$ (set $\delta = 0$ in $\frac{\partial c_m}{\partial n_m^c} = \delta n_m^o + n_m^c$). Profit maximization implies that each censored-content mass-mail sender equates marginal revenue with marginal cost:

$$p^o = \delta n_m^c + n_m^o \tag{68}$$

and

$$\frac{p^c}{d_m^{c,\infty}\gamma^c} = n_m^{c,\infty} \tag{69}$$

Since we have already assumed that $p^o \geq p^c$ and $d^{c,\infty}\gamma^c \geq 1$, we have

$$p^o \geq \frac{p^c}{d^{c,\infty}\gamma^c} \tag{70}$$

Substitute the value of the above inequalities from the two profit maximization conditions above, we have:

$$\delta n_m^c + n_m^o \geq n_m^{c,\infty} \iff \tag{71}$$

$$n_m^o \geq n_m^{c,\infty} \text{ because } \delta < 0 \implies \tag{72}$$

$$ctt_o \times n_m^o \geq ctt_c \times n_m^{c,\infty} \text{ if } ctt_o \geq ctt_c \geq 0 \implies \tag{73}$$

$$n_r^o \geq n_r^{c,\infty} \tag{74}$$

The last inequality holds because of the following. We already assumed that the volume of censored-content targeted-mail is negligible[20]. Then we could interpret the constants ctt_o and ctt_c as the reciprocals of the

[20] Alternatively, by properly defining the costs function of the censored-content targeted-mail senders, we could assume a weaker condition that the ratio of the volume of mail sent by censored-content targeted- over mass-mail senders in the open channel is greater or equal to the corresponding ratio in the censored channel.

number of recipients of censored-content mass-mail in the open channel and censored channels, respectively. But by definition the censored channel is essential, there must be at least or greater number of recipients in the censored channel. So $ctt_o \geq ctt_c$. Since mass-mail will be distributed evenly, we have $ctt_o \times n_m^o = n_r^o$, and $ctt_c \times n_m^{c,\infty} = n_r^{c,\infty}$. Hence $n_r^o \geq n_r^{c,\infty}$.

(ii) Inequality (b) is $u^r_{\text{Type 1 errors}} \leq \lim_{\gamma^o, \hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 1 errors}}$, or equivalently:

$$\begin{aligned} &\sum_{j \in \{o,c\}} \epsilon \kappa_r^j \frac{\hat{n}_r^j}{\hat{\gamma}^j} + \sum_{j \in \{o,c\}} (1-\phi_t^j) \kappa_r^j \frac{b(d^j) n_r^j}{d^j \gamma^j} \\ &\leq \epsilon \frac{\kappa_r^c \hat{n}_r^{c,\infty}}{\hat{\gamma}^c} + \frac{(1-\phi_t^c) \kappa_r^c b(d^{c,\infty}) n_r^{c,\infty}}{d^{c,\infty} \gamma^c} \end{aligned} \tag{75}$$

Similarly, by the implication of the assumption of no uncensored mail in the open channel, we have $\epsilon \frac{\kappa_r^c \hat{n}_r^c}{\hat{\gamma}^c} = \epsilon \frac{\kappa_r^c \hat{n}_r^{c,\infty}}{\hat{\gamma}^c}$, and $\epsilon \frac{\kappa_r^o \hat{n}_r^o}{\hat{\gamma}^o} = 0$. It is equivalent to proving that:

$$\sum_{j \in \{o,c\}} (1-\phi_t^j) \kappa_r^j \frac{b(d^j) n_r^j}{d^j \gamma^j} \leq \frac{(1-\phi_t^c) \kappa_r^c b(d^{c,\infty}) n_r^{c,\infty}}{d^{c,\infty} \gamma^c} \tag{76}$$

The above is true for $\phi_t^o = 1$ because both sides are zero. For $\phi_t^o = 0$, because $\kappa_r^c = 1$ and $\kappa_r^o = 0$ by Proposition 1, it is equivalent to proving that:

$$\frac{b(d^c) n_r^c}{d^c} \leq \frac{b(d^{c,\infty}) n_r^{c,\infty}}{d^{c,\infty}} \tag{77}$$

To prove that $\frac{b(d^c) n_r^c}{d^c} \leq \frac{b(d^{c,\infty}) n_r^{c,\infty}}{d^{c,\infty}}$, as before, profit maximization implies that:

$$\frac{p^c}{d_m^{c,\infty} \gamma_m^c} = n_m^{c,\infty} \tag{78}$$

$$\frac{p^c}{d_m^c \gamma_m^c} = n_m^c \tag{79}$$

Together we have:

$$n_m^{c,\infty} d_m^{c,\infty} = n_m^c d_m^c \tag{80}$$

Again assuming that the volume of censored-content targeted-mail is negligible, the above equation implies:

$$n_s^{c,\infty} d^{c,\infty} = n_s^c d^c \tag{81}$$

$$n_r^{c,\infty} d^{c,\infty} = n_r^c d^c \text{ by even distribution of mass mail} \tag{82}$$

Now substitute $n_r^{c,\infty} = n_r^c d^c / d^{c,\infty}$ into $\frac{b(d^c)n_r^c}{d^c} \leq \frac{b(d^{c,\infty})n_r^{c,\infty}}{d^{c,\infty}}$:

$$\frac{b(d^c)}{(d^c)^2} \leq \frac{b(d^{c,\infty})}{(d^{c,\infty})^2} \tag{83}$$

This inequality is guaranteed because: (1) we already assumed that $b(d^j) \equiv (d^j)^2$ (in fact, we need a weaker condition that $b(\cdot)$ is increasing at a smaller (or equal) rate than that of the reciprocal of the square of its arguments. (2) $d^c \geq d^{c,\infty}$. To prove $d^c \geq d^{c,\infty}$, note that in Proposition 2, the disguise level is a constant in cases (a) and (b), so $d^c = d^{c,\infty}$. In case (c), $d_m^c = \frac{(p^c)^2}{(1-\delta^2)(\gamma^c)^2 + \delta p^o p^c \gamma^c}$. We will just rename $d_m^c = d^c$ because there is no censored-content targeted-mail by assumption. We can obtain $d^{c,\infty} = \frac{(p^c)^2}{(\gamma^c)^2}$ by setting $\delta = 0$ in d_s^c. So $d_s^c \geq d^{c,\infty}$ if and only if

$$\frac{(p^c)^2}{(1-\delta^2)(\gamma^c)^2 + \delta p^o p^c \gamma^c} \geq \frac{(p^c)^2}{(\gamma^c)^2} \iff \tag{84}$$

$$(\gamma^c)^2 \geq (1-\delta^2)(\gamma^c)^2 + \delta p^o p^c \gamma^c \iff \tag{85}$$

$$\delta^2 (\gamma^c)^2 \geq \delta p^o p^c \gamma^c \tag{86}$$

The last inequality is true because $\delta < 0$.

(iii) Inequality (c) is $u^r_{\text{Type 2 errors}} \leq \lim_{\gamma^o, \hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 2 errors}}$, or equivalently:

$$\sum_{j \in \{o,c\}} (1-\epsilon)\kappa_r^j \hat{n}_r^j (1 - \frac{1}{\hat{\gamma}^j}) + \sum_{j \in \{o,c\}} \phi_t^j \kappa_r^j b(d^j) n_r^j (1 - \frac{1}{d^j \gamma^j})$$
$$\leq (1-\epsilon)\kappa_r^c \hat{n}_r^{c,\infty} (1 - \frac{1}{\hat{\gamma}^c}) + \phi_t^c \kappa_r^c b(d^{c,\infty}) n_r^{c,\infty} (1 - \frac{1}{d^{c,\infty}\gamma^c}) \tag{87}$$

Similarly, by the profit-maximizing implication of the assumption that there is no uncensored-content mail in the open channel, we have $(1-\epsilon)\kappa_r^c \hat{n}_r^c (1 - \frac{1}{\hat{\gamma}^c}) = (1-\epsilon)\kappa_r^c \hat{n}_r^{c,\infty} (1 - \frac{1}{\hat{\gamma}^c})$, and since $(1-\epsilon)\kappa_r^o \hat{n}_r^o (1 - \frac{1}{\hat{\gamma}^o}) = 0$ because $\hat{\gamma}^o = 1$. It is equivalent to proving that:

$$\sum_{j \in \{o,c\}} \phi_t^j \kappa_r^j b(d^j) n_r^j (1 - \frac{1}{d^j \gamma^j}) \leq \phi_t^c \kappa_r^c b(d^{c,\infty}) n_r^{c,\infty} (1 - \frac{1}{d^{c,\infty}\gamma^c}) \tag{88}$$

The above is true for $\phi_t^o = 0$ because both sides are zero. For $\phi_t^o = 1$, because $\kappa_r^c = 1$ and $\kappa_r^o = 1$ by Proposition 1, it is equivalent to proving that:

$$\sum_{j=o,c} b(d_s^j) n_r^j (1 - \frac{1}{d^j \gamma^j}) \leq b(d^{c,\infty}) n_r^{c,\infty} (1 - \frac{1}{d^{c,\infty}\gamma^c}) \tag{89}$$

But there is no type 2 errors with censored-content mail in the open channel because $\gamma^o = 1$ (implying $d^o = b(d^o) = 1$), so it is equivalent to proving that:

$$b(d^c)n_r^c(1 - \frac{1}{d^c\gamma^c}) \leq b(d^{c,\infty})n_r^{c,\infty}(1 - \frac{1}{d^{c,\infty}\gamma^c}) \tag{90}$$

Substitute $n_r^{c,\infty} = n_r^c d_s^c / d^{c,\infty}$ and cancelling n_r^c on both sides:

$$b(d^c)d^{c,\infty}(1 - \frac{1}{d^c\gamma^c}) \leq b(d^{c,\infty})d^c(1 - \frac{1}{d^{c,\infty}\gamma^c}) \tag{91}$$

As a special case, substitute $b(d^j) \equiv (d^j)^2$:

$$(d^c)^2 d^{c,\infty}(1 - \frac{1}{d^c\gamma^c}) \leq (d^{c,\infty})^2 d_s^c(1 - \frac{1}{d^{c,\infty}\gamma^c}) \iff \tag{92}$$

$$d^c(1 - \frac{1}{d^c\gamma^c}) \leq d^{c,\infty}(1 - \frac{1}{d^{c,\infty}\gamma^c}) \iff \tag{93}$$

$$d^c \leq d^{c,\infty} \tag{94}$$

To show that $d^c \leq d^{c,\infty}$, we actually need to show $d^c = d^{c,\infty}$ because we already know that $d^c \geq d^{c,\infty}$ in order to satisfy inequality (b). Now already know from the proof of inequality (b) that $d^c = d^{c,\infty}$ in cases (a) and (b) in Proposition 2, and for case (c) $d^c \leq d^{c,\infty}$ if $\delta^2(\gamma^c)^2 \leq \delta p^o p^c \gamma^c$ or either p^o or p^c equal to zero.

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.

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`.

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`.

Sophos. *Sophos Security Threat Management Report*. 2005.

Timothy van Zandt. Information overload in a network of targeted communication. *RAND Journal of Economics*, 35(3):542–560, 2004.