Working Paper #06-20

October 2006

Using Uncensored Communication Channels to Divert Spam Traffic

Benjamin Chiao and Jeffrey MacKie-Mason
University of Michigan

Using Uncensored Communication Channels to Divert Spam Traffic*

Benjamin Chiao† and Jeffrey MacKie-Mason‡ 8/31/06 (Revised: 10/15/06)
{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. We adopt an incentive-centered design approach to develop a simple, feasible improvement to the current email system using 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 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

*We appreciate comments from Nat Bulkley, Zhuoran Chen, Nick Economides, Michael Hess, Peter Honeyman, Paul Resnick, Doug Van Houweling, Michigan China Fellows, 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 and at the Telecommunications Policy Research Conference. 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

for all recipients the satisfaction associated with desirable mail received increases, and 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 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 (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 mailing list, which one could unsubscribe from.

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 re-generated, 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 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 should be real. No buyers will believe that a $50 Rolex is authentic.

Casual 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. We refer to these as "censored" commercial solicitations. Such "censorship", incarnated as filters and domain-blocking rules, are ubiquitous at the email service provider' level. Of course, the "censorship" of which we speak is not necessarily 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 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 expected benefit to them exceeds the expected 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 (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

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

[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.

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.[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 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] Note that email from large legitimate businesses such as eBay and IBM is not considered spam in our earlier definition because it is only initially (but not persistently) unsolicited. Though such email could be quite bothersome, especially if one needs to receive a small portion of wanted email from such businesses.

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 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.

[3]In 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.

[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 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. 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.).

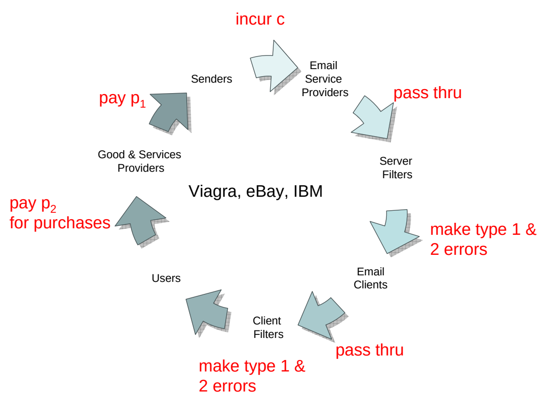

Figure 1: Stakeholders in an e-mail ecosystem.

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 advertising[5]: 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 un-

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

Figure 2: An hypothetical open channel.

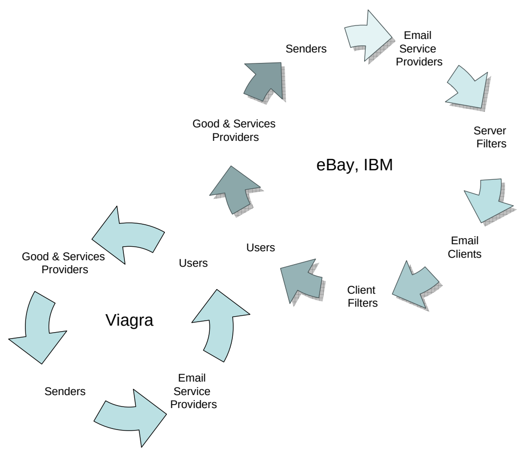

Figure 3: Separating the demand for and supply of bulk unsolicited commercial advertising.

der 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][6], challenge-response [Dwork and Naor, 1993, Laurie and Clayton,

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

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 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.[7] 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.

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 spammers to deliver advertisements to email recipients, some of whom in turn willingly choose to purchase the advertised products[8]. 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.

[7] 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 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.

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

3.1 Mail Types

Mail types coincide with senders' types. Such types are defined by two attributes: mass or targeted mail, and censored or uncensored content.[9]

The first attribute is mainly a cost attribute. The content creation cost per copy of mass mail is much lower than that of targeted mail.[10] 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 whether the sender sends content of a type that is censored (if recognized) by the email service provider.[11,12]

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-mail senders to reduce the supply

[9]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.

[10]We 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.

[11]Recipient censorship (with, for example, personal spam filters) is 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.

[12]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 practice, much spam can be automatically identified as being sent from a censored-content mass-mail sender, but our results are robust as long as considerable spam cannot.

of their messages delivered to the current standard email channel (the *censored* channel). Therefore, we simplify by making this assumption:

Assumption 1 *Mass-mail senders send only censored content, and targeted-mail senders send only uncensored content.*[13]

3.2 The Recipients' Problem

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:

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

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

Channel j is either (o)pen or (c)ensored. 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 $t \in \{(h)\text{igh}, (l)\text{ow}\}$ desires censored-content mail in channel j, then $\phi_t^j = 1$; otherwise $\phi_t^j = 0$.[14] 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:

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

[13]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.

[14]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 do not want any. Thus, there is the possibility of both Type 1 and Type 2 errors for each.

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

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

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

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

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 $\gamma^c \geq \hat{\gamma}^c$. 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, $\gamma^o = 1$.[15]

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, $d_s^j \in [\frac{1}{\gamma^j}, 1]$, for mail sent to channel j, where d_s^j is a multiplicative factor adjusting the 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; we assume that there is no effort made to disguise content in the open channel (by definition of the lower bound of d_s^j, $d_s^o = 1$ because $\gamma^o = 1$ implies that the upper and lower bounds coincide).

Earlier on, we assumed that censored-content mail is sent only by mass-mail senders, who evenly distribute such mail to all recipients. We therefore define $n_r^j = \frac{\sum_m n_m^j}{R^j}$ as the volume of censored-content mail sent to recipient r in channel j, where n_m^j is the censored-content mass-mail sent by sender m to channel j, and R^j is the number of recipients using channel j. Then the portion that actually reaches recipient r is $\frac{n_r^j}{d^j\gamma^j}$, where d^j is the weighted average of disguise levels. Similarly, except of course that targeted mail does not have to be averaged out across recipients, we use the hat symbol to denote the corresponding uncensored-content variables: $\hat{n}_r^j, \hat{d}^j$ and $\hat{\gamma}^j$. Since we assume that there is no need to disguise uncensored content, $\hat{d}^j = 1$.

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

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 information preference function, which is increasing in the effort made to disguise censored-content mail. To allow for more generality that will be clear shortly, we specify that this function can be different for different mail. Namely, the functions are $I(d^j)$ and $\bar{I}(d^j)$ for filtered and unfiltered 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:

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

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 to facilitate purchases).

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

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

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 suffer from receiving it).

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

$$u^r_{\text{Type 2 errors}} = \sum_{j\in\{o,c\}} (1-\epsilon)\kappa_r^j \hat{n}_r^j (1-\frac{1}{\hat{\gamma}^j}) + \sum_{j\in\{o,c\}} \phi_t^j \kappa_r^j \bar{I}(d^j) n_r^j (1-\frac{1}{d^j \gamma^j}) \tag{7}$$

where the first term is desired filtered uncensored mail, and the second term is filtered censored-content mail for high type recipients.[16]

To simplify the model, we rule out the unlikely scenario that no one is using the existing email channel:

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

We also make another assumption that will greatly simplifying the notations and algebra:

Assumption 3 *There is no uncensored-content mail in the open channel. That is,* $\hat{n}_r^o = 0$.

This assumption is justified in two senses. First, when there is less spam in the censored channel, uncensored-content senders will not want to use the open channel if the risk of it being filtered out in the open channel is getting less. Second, uncensored-content mail will be mingled with possibly much censored-content mail, uncensored-content senders will need to weigh the chances of the mail being filtered in the censored channel and being mingled in the open channel, especially if sending to an additional channel (even without filtering as in the open channel) incurs extra costs.

Assumption 4 *The information preference functions take the following form:* $I(d^j) = (d^j)^{\beta^r}$ *and* $\bar{I}(d^j) = (d^j)^{\bar{\beta}^r}$.

The basic idea of the information preference function is to specify how an individual makes the tradeoff between information and volume. The above assumption implies that the (dis-)utility associated with unfiltered censored-content mail, which is $\frac{I(d^j)n_r^j}{d^j\gamma^j}$, becomes $(d^j)^{\beta^r-1}\frac{n_r^j}{\gamma^j}$. For $\beta^r = 0$, such (dis-)utility becomes $\frac{n_r^j}{d^j\gamma^j}$. Recipient r cares about the effective unfiltered volume not the information content. For $\beta^r = 1$, such

[16]One could elaborate by allowing Type 2 errors associated with targeted mail to be more annoying. This is because mass mail always appears in multiple and sometimes almost identical copies in a given recipient's inbox. This higher substitutability implies that there is *a low Type 2 error cost* associated with mass mail. That is, one could redefine spam as mail with lower Type 2 error cost. In other words, 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.

Mathematically, one could define the first term to increase at a higher rate than the second term in $u^r_{\text{Type 2 errors}}$.

(dis-)utility becomes $\frac{n_r^j}{\gamma^j}$, recipient r is indifferent or neutral to any disguise levels associated with unfiltered censored-content mail. For $\beta^r \in (0,1), (d^j)^{\beta^r-1} \geq 1,$ such (dis-)utility is weakly greater than $\frac{n_r^j}{\gamma^j}$. That is, conditional on the same volume of mail sent (n_r^j), high (low) type recipients would rather (not) have a higher volume of disguised mail unfiltered than a lower volume (since γ^j is not adjusted by the disguise level that equals 1) of undisguised mail unfiltered. For $\beta^r > 1$, such (dis-)utility is weakly less than $\frac{n_r^j}{\gamma^j}$. That is, conditional on the same n_r^j, high (low) type recipients would rather (not) have a lower volume of undisguised mail unfiltered than a higher volume of disguised mail unfiltered.

Similarly, the information neutral cuttoff for the dis-utility associated with filtered censored-content mail is $\bar{\beta}^r = \bar{\beta}^{*r} \equiv \log(\frac{1-\frac{1}{\gamma^j}}{1-\frac{1}{d^j\gamma^j}})/\log d^j$.[17]

If the comparison to the "information neutral" benchmark cases offers a useful tool to understanding the tradeoff between volume and information, having a different function for filtered and unfiltered mail, just as what we did using $I(d^j)$ and $\bar{I}(d^j)$, allows a consistent description of a given individual's preference for information across mail types. For example, a theorist could require that a given individual must be either one of the following: $\{(\beta^r = 1$ and $\bar{\beta}^r = \bar{\beta}^{*r}), (\beta^r < 1$ and $\bar{\beta}^r < \bar{\beta}^{*r}), (\beta^r > 1$ and $\bar{\beta}^r > \bar{\beta}^{*r})\}$. On the other hand, one could also not rely on such triplet to leave room for behavioral inconsistencies.

As a simplification, we assume that the information preference functions are homogenous. That is, $\forall r$, $\beta^r = \beta$,and $\bar{\beta}^r = \bar{\beta}$.

Let us now state the recipient's problem formally. Given the filter strengths, disguise levels, email volume and actions of other recipients,

[17] Note that only high type recipients experience the dis-utility of $\bar{I}(d^j)n_r^j(1 - \frac{1}{d^j\gamma^j}) = (d^j)^{\bar{\beta}^r} n_r^j - (d^j)^{\bar{\beta}^r-1}\frac{n_r^j}{\gamma^j}$. For $\bar{\beta}^r = 0$, such dis-utility becomes $n_r^j - \frac{n_r^j}{d^j\gamma^j}$. High type recipients suffer from the effective filtered volume not the information content. For $\bar{\beta}^r = \bar{\beta}^{*r} \equiv \log(\frac{1-\frac{1}{\gamma^j}}{1-\frac{1}{d^j\gamma^j}})/\log d^j$, such dis-utility becomes $n_r^j - \frac{n_r^j}{\gamma^j}$, high type recipients are indifferent to any disguise levels associated with filtered censored-content mail. To find $\bar{\beta}^{*r}$,equate $(d^j)^{\bar{\beta}^r} n_r^j - (d^j)^{\bar{\beta}^r-1}\frac{n_r^j}{\gamma^j}$ with $n_r^j - \frac{n_r^j}{\gamma^j}$ getting:

$$(d^j)^{\bar{\beta}^r}(1 - \frac{1}{d^j\gamma^j}) = 1 - \frac{1}{\gamma^j} \iff \bar{\beta}^r \log d^j = \log(\frac{1-\frac{1}{\gamma^j}}{1-\frac{1}{d^j\gamma^j}}) \iff \bar{\beta}^{*r} = \frac{\log(\frac{1-\frac{1}{\gamma^j}}{1-\frac{1}{d^j\gamma^j}})}{\log d^j}$$

For $\bar{\beta}^r \in (0, \bar{\beta}^{*r})$, such dis-utility is less than $n_r^j - \frac{n_r^j}{\gamma^j}$. That is, conditional on the same n_r^j, high type recipients would suffer less with a smaller volume of disguised mail filtered than a higher volume of undisguised mail filtered. For $\beta^r > \bar{\beta}^{*r}$, such dis-utility is more than $n_r^j - \frac{n_r^j}{\gamma^j}$. That is, conditional on the same n_r^j, high type recipients would suffer more with a smaller volume of disguised mail filtered than a higher volume of undisguised mail filtered.

recipient r makes a binary choice of whether to read mail in the open channel, $\kappa_r^o \in \{0, 1\}$, 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 *With Assumptions 1, 2, and 3, 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 ($\phi_l^j = 0$), reading the open channel provides no benefit, but creates dis-utility 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 1 errors (6). Likewise it has no effect on Type 2 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 behavior of spammers when an open channel is introduced, after which we analyze the advertisers' problem, and the welfare effects of an open channel.

3.3 The Senders' Problem

We will describe in detail the cost and revenue functions of the censored-content mass-mail senders only. Again, 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-mail 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 $\partial c_m / \partial n_m^j > 0$.[18] We allow for economies of scale in the sense of sub-additivity,

[18]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].

$c_m(n^o_m, 0, d^c_m) + c_m(0, n^o_m, d^c_m) > c_m(n^o_m, n^c_m, d^c_m)$, and cost complementarity (ie., $\frac{\partial^2 c_m}{\partial n^j_m \partial n^i_m} < 0, i \neq j$). To be concrete, we specify $c_m(n^o_m, n^c_m, d^c_m) = FC_m + g_m(d^c_m) + \delta n^o_m n^c_m + \frac{1}{2}(n^o_m)^2 + \frac{1}{2}(n^c_m)^2$, in which $g_m(d^c_m) = \frac{1}{d^c_m} - 1$[19], so that the cost of no disguising ($d^c_m = 1$) is $g_m(1) = 0$. Cost complementarity and subadditivity are both ensured by letting $\delta < 0$.[20] We also assume a regularity condition of $\delta^2 < 1$.

On the revenue side, senders are price takers. Sellers of censored goods pay them for solicitations. Let p^j be the advertising charge per disguised email $(\frac{n^j_m}{d^j_m \gamma^j})$ reaching the users in channel j[21].

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 censored channel, the sender does not tag the message, and in fact may expend some effort to disguise the content. We assume that mass mail sent 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.

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

Next we state the solutions to the above maximization problem:

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

[20] Cost complementarity follows from $\delta < 0$ because $\frac{\partial c_m}{\partial n^c_m} = \delta n^o_m + n^c_m$, and $\frac{\partial c_m}{\partial n^o_m} = \delta n^c_m + n^o_m$. Subadditivty does as well because

$$\begin{aligned}
&c_m(n^o_m, n^c_m, d^c_m) - [c_m(n^o_m, 0, d^c_m) + c_m(0, n^c_m, d^c_m)] \\
=&FC_m + g_m(d^c_m) + \delta n^o_m n^c_m + \frac{1}{2}(n^o_m)^2 + \frac{1}{2}(n^c_m)^2 - [FC_m + \\
&\frac{1}{2}(n^o_m)^2 + FC_m + g_m(d^c_m) + \frac{1}{2}(n^c_m)^2] \\
&= \delta n^o_m n^c_m - FC_m < 0.
\end{aligned}$$

[21] 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 $\frac{n^j_m}{d^j_m \gamma^j}$) 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.

Proposition 2 *The best responses of sender m are*[22]*:*
Case (a), $p^o \leq \frac{p^c}{\delta\gamma^c} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$:

$$d_m^{*c} = 1; n_m^{*o} = \frac{1}{1-\delta^2}(p^o - \frac{\delta p^c}{\gamma^c}); n_m^{*c} = \frac{1}{1-\delta^2}(\frac{p^c}{\gamma^c} - \delta p^o) \qquad (11)$$

Case (b), $p^o \geq \frac{p^c}{\delta} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$:

$$d_m^{*c} = \frac{1}{\gamma^c}; n_m^{*o} = \frac{1}{1-\delta^2}(p^o - \delta p^c); n_m^{*c} = \frac{1}{1-\delta^2}(p^c - \delta p^o) \qquad (12)$$

We next discuss the implications of Proposition 2.

Mass mail sent to a particular channel is increasing in the advertising price of that channel, and of the other channel only when there is complementarity ($\delta \neq 0$). Any possible increase is at a lower rate when the complementarity is weak (i.e., $|\delta|$ is small). In fact, when complementarity is weaker, mass mail sent will be lower in both channels. Also, as long as either p^o or p^c is (or both are) strictly positive, there will be mass-mail sent to both channels unless $\delta = 0$. In other words, spam will not be totally eliminated when $\delta \neq 0$. When there is complementarity, it is optimal to send some positive volume to the channel even though the price there is zero. This is because this helps reduce the volume generation cost in the other channel.

Notice that the marginal revenues of sending n_m^{*o} and $\frac{n_m^{*c}}{d_m^{*c}\gamma^c}$ are p^o and p^c. Alternatively, one could regard the marginal revenues of each n_m^{*o} and $\frac{n_m^{*c}}{d_m^{*c}}$ as p^o and $\frac{p^c}{\gamma^c}$. To increase the total revenue in the censored channel by the same amount, a sender could either adjust n_m^{*c} or d_m^{*c}, depending on which is cheaper. Since the reciprocal of the disguise cost is linear and the volume cost is convex, if they are on the same plane, they intersect at one point. Left of this point, the volume cost is less convex, so the marginal volume cost is less than the marginal disguise cost. Adjusting volume but not disguising is cheaper. Right of this point, the opposite happens. Disguising is cheaper so one will never increase n_m^{*c} beyond the intersection point.

The intuition for the effects of γ^c is the following. In case (a), d_m^{*c} is unchanged from the default level. The only degrees of freedom to adjusting revenue are n_m^{*o} and n_m^{*c}. When γ^c increases, $\frac{p^c}{\gamma^c}$ decreases. Before the change of γ^c, profit maximization must implies that the marginal

[22]For the trivial case of $p^o = p^c = 0$, we have shown in Appendix 6.1 that:

$$d_m^{*c} = 1; n_m^{*o} = 0; n_m^{*c} = 0.$$

revenue $\frac{p^c}{\gamma^c}$ must equal some marginal cost. Thus, when $\frac{p^c}{\gamma^c}$ decreases, it must equal some lower marginal cost. But since the total cost is convex, n_m^{*c} must be lowered conditional on the same or smaller n_m^{*o}. This can be confirmed from the best response of n_m^{*c} in case (a), where both n_m^{*c} and n_m^{*o} are decreasing in γ^c. n_m^{*o} is decreasing in γ^c for a different reason. Now we knew that an increase in γ^c decreases n_m^{*c}. But the marginal cost of sending n_m^o is $\frac{\partial c_m}{\partial n_m^o} = \delta n_m^c + n_m^o$, reducing n_m^{*c} increases $\frac{\partial c_m}{\partial n_m^o}$ since $\delta < 0$. Again, by equating the marginal revenue p^o with some marginal cost, n_m^{*o} must be lowered. More generally, γ^c is negatively related to n_m^{*o} and n_m^{*c}. In case (b), d_m^{*c} varies inversely with γ^c. In fact, d_m^{*c} is at the lowest bound of $\frac{1}{\gamma^c}$. In this case, only d_m^{*c} adjusts because reducing d_m^{*c} is always cheaper than increasing n_m^{*c} when γ^c changes. Again, since n_m^{*o} and d_m^{*c} are only related through n_m^c in $\frac{\partial c_m}{\partial n_m^o}$, when n_m^{*c} is unchanged, so is n_m^{*o}.

It is important to emphasize again that the volumes sent are not necessarily the same as the volumes received. The latter is what the recipients should be concerned about in addition to the disguise levels. We defer the utility changes due to disguise levels until the welfare analysis in the next section. With some intuition explained above, we are ready to summarized the properties of volumes sent and received in the following proposition:

Proposition 3 *Consider the two cases in Proposition 2.*
(1) In the censored channel, the total volume of email received is: i) increasing in both p^o and p^c; ii) decreasing in δ; iii) decreasing in γ^c in case a, and independent in case b. The total volume of email sent is: i) increasing in p^o and p^c, decreasing in p^c; ii) unchanged in δ; iii) decreasing, and unchanged in γ^c in cases a and b, respectively.
(2) In the open channel, the total volume of email received vis-a-vis sent are the same by construction (i.e., $\frac{n_m^{*o}}{d_m^{*o}\gamma^o} = n_m^{*o}$). Such volumes are: i) increasing in p^o and p^c; iii) decreasing in δ; iv) decreasing, and unchanged in γ^c in cases a and b, respectively.

Proof. In the censored channel, the total volume of email received are:
Case (a), $p^o \leq \frac{p^c}{\delta\gamma^c} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$:

$$\frac{n_m^{*c}}{d_m^{*c}\gamma^c} = \frac{1}{1-\delta^2}[\frac{p^c}{(\gamma^c)^2} - \frac{\delta p^o}{\gamma^c}] \tag{13}$$

Case (b), $p^o \geq \frac{p^c}{\delta} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$:

$$\frac{n_m^{*c}}{d_m^{*c}\gamma^c} = n_m^{*c} = \frac{1}{1-\delta^2}(p^c - \delta p^o) \tag{14}$$

All directions of change here can be seen by inspection or straightforward differentiations. ■

3.4 The Advertisers' Problem

Advertiser a pays an advertising charge of p^j for each email message that passes through the filter in channel j. For each product or service sold, the advertiser collects a sales revenue of s. The probability of purchase is $\theta^j \equiv \theta(\frac{R_{\phi_t^j=1}}{R^j}, d_a^j - d_a^i)$ for each p^j paid, where $R_{\phi_t^j=1}$ is the number of high type recipients in channel j, and $d_a^j \in [\frac{1}{\gamma^j}, 1]$ is the disguise level associated with n_a^j, which is the volume of email advertiser a asks the spammers to successfully pass through the filter in channel j.[23] $\frac{R_{\phi_t^j=1}}{R^j}$ measures the average ϕ_t^j in channel j.$d_a^j - d_a^i$ measures the informativeness of the messages relative to the counterpart in the other channel. $\theta_1 > 0$ because of a higher propensity to purchase is tautologically associated with a higher average ϕ_t^j. $\theta_2 > 0$ because of an assumption that a greater informativeness of the messages relative to the counterpart in the other channel is assumed to facilitate purchase better.

The cost function consists of two parts: production and advertising. The total production cost, $k(\sum_j \theta^j n_a^j)$, is a convex function as usual, where $\sum_j \theta^j n_a^j$ is the total quantity of goods or services sold. The total advertising cost, $\sum_j p^j n_a^j$, is linear however. The spammer's market is competitive. Spammers as price takers treats the marginal revenue of spamming, p^j, as given. For such market to clear, p^j must also equal to the marginal advertising cost for the advertisers. Hence, the total advertising cost is linear. In other words, it only chooses $\{n_a^j\}_{j=c,o}$ to maximize expected profits because $\{n_a^j\}_{j=c,o}$ is sufficiently small among the many advertisers that its advertising demand does not drive up or down the marginal advertising cost p^j. That is, p^j is independent of $\{n_a^j\}_{j=c,o}$. Of course, the aggregate level $\{\sum_a n_a^j\}_{j=c,o}$ will affect p^j but we will ignore this effect.

Mathematically, advertiser a chooses $\{n_a^j\}_{j=c,o}$ to maximize:

[23]Note that θ is not a function of the volume of email advertiser a asks the spammers to successfully pass through the filter in each channel. It is reasonable in the sense that recipients are more likely to make a purchase if they are being repeatedly reminded of, and that such volume is sufficient small so that each recipient is not likely to receive multiple email by the same advertiser. We therefore rule out the case in which the diminishing returns kick in a significant way when there are multiple advertisers selling the same services or products.

$$\pi_a = \sum_j [s\theta(\frac{R_{\phi_t^j=1}}{R^j}, d_a^j - d_a^i) - p^j] n_a^j - k(\sum_j \theta^j n_a^j) \tag{15}$$

The advertisers' maximization problem gives the marginal revenue equals marginal cost condition:

$$\forall j \neq i : p^j = \theta^j [s - k'(\sum_j \theta^j n_a^j)] \tag{16}$$

Now we will show the relationships between advertising charges in each channel, before and after the implementation of the open channel. (Note that we use ∞ to denote variables when the open channel is absent. When the open channel is absent, we define it by saying that $\delta = 0$, and $p^o = 0$ [or equivalently $\gamma^o, \hat{\gamma}^o \to \infty$]. Denote $\theta^\infty(\frac{R_{\phi_t^c=1}}{R^c}, d_a^{c,\infty})$ as the response function when the open channel is absent.)

Proposition 4 *(i) The advertising charge in the open channel is weakly higher than that in the censored channel (and than that in the censored channel when the open channel is absent if a sufficient condition is met). (ii) The advertising charge in the censored channel weakly decreases when the open channel is implemented. (iii) The advertising charges are increasing in the average propensity to purchase.*

Proof. (i) By (16), $p^o - p^c = (\theta^o - \theta^c)[s - k'(\sum_j \theta^j n_a^j)]$. Again by (16), $[s - k'(\sum_j \theta^j n_a^j)]$ is positive for positive prices. We just need to show that $(\theta^o - \theta^c) \geq 0$. By Proposition 1, $\frac{R_{\phi_t^o=1}}{R^o} = \frac{R_{\phi_t^o=1}}{R_{\phi_t^o=1}} \geq \frac{R_{\phi_t^c=1}}{R^c}$. Also, by construction, $d_a^o = 1 \geq d_a^c$. So $d_a^o - d_a^c \geq 0$, and $d_a^c - d_a^o \leq 0$. These imply that $\theta^o \equiv \theta(\frac{R_{\phi_t^o=1}}{R^o}, d_a^o - d_a^c) \geq \theta(\frac{R_{\phi_t^c=1}}{R^c}, d_a^c - d_a^o) \equiv \theta^c$. Hence, we have $p^o \geq p^c$. To have $p^o \geq p^{c,\infty}$, by (16), we require a sufficient condition of $\theta^o \geq \theta^{c,\infty}$ or equivalently:

$$\theta(1, 1 - d_a^c) \geq \theta^\infty(\frac{R_{\phi_t^c=1}}{R^c}, d_a^{c,\infty}). \tag{17}$$

(ii) First, by Proposition 1, $\frac{R_{\phi_t^c=1}}{R^c}$ remains the same when the open channel is implemented because the high type recipients only use the open channel additionally while continue using the censored channel. Second, $d_a^c - d_a^o = d_a^c - 1 \leq 0$, so $d_a^{c,\infty} \geq 0 \geq d_a^c - d_a^o$. These imply that $\theta^\infty(\frac{R_{\phi_t^c=1}}{R^c}, d_a^{c,\infty}) \geq \theta(\frac{R_{\phi_t^c=1}}{R^c}, d_a^c - d_a^o)$. We have $p^{c,\infty} \geq p^c$ by (16) because $[s - k'(\sum_j \theta^j n_a^j)]$ and $[s - k^{\infty\prime}(\theta^{c,\infty} n_a^c)]$ are positive for positive prices.

(iii) Taking a derivative of (16) w.r.t. $\frac{R_{\phi_t^j=1}}{R^j}$:

$$\frac{\partial p^j}{\partial(\frac{R_{\phi_t^j=1}}{R^j})} = \theta_1^j[s - k'(\sum_j \theta^j n_a^j)] + \theta^j \frac{\partial k'(\sum_j \theta^j n_a^j)}{\partial(\frac{R_{\phi_t^j=1}}{R^j})}$$
$$= \theta_1^j[s - k'(\sum_j \theta^j n_a^j)] + \theta^j k'' n_a^j \theta_1^j \geq 0$$

■

To solve for the sufficient condition in Proposition 4, one needs to specify a functional form of the response function.

Assumption 5 *The response function takes the form of an exponential distribution. More specifically,* $\theta(\frac{R_{\phi_t^j=1}}{R^j}, d_a^j - d_a^i) \equiv 1 - e^{-(\frac{R_{\phi_t^j=1}}{R^j} + e^{d_a^j - d_a^i})}$.

The sufficient condition takes a simple form under this exponential distribution, which is $1 - d_a^c \geq d_a^{c,\infty}$.[24]

3.5 Welfare

Proposition 5 *Under a sufficient condition, the implementation of the open channel decreases the volume of censored-content mass mail unfiltered in or sent to the censored channel and its associated dis-utility for recipients who do not want it.*

Proof. We have showed that $n_m^{*c,\infty} \geq n_m^{*c}$, $\frac{n_m^{*c,\infty}}{d_m^{*c,\infty}\gamma^c} \geq \frac{n_m^{*c}}{d_m^{*c}\gamma^c}$ and $\frac{I(d^{*c,\infty})n_r^{*c,\infty}}{d^{*c,\infty}\gamma^c} \geq \frac{I(d^{*c})n_r^{*c}}{d^{*c}\gamma^c}$ in Appendix 6.2 under some sufficient conditions. ■

Whether such sufficient condition will be satisfied depends on the values of the parameters and further specifying some functional forms, which will necessarily restrict the generality of our results. We will leave it as it is but instead offer some general intuition about when the sufficient condition is more likely to satisfy. It is true that the introduction of the open channel decreases the probability of a sale in the censored channel because $\theta(\frac{R_{\phi_t^c=1}}{R^c}, d_a^c - d_a^o) \leq \theta(\frac{R_{\phi_t^c=1}}{R^c}, d_a^{c,\infty})$. Subsequently, $p^{c,\infty} \geq p^c$ as we have seen in Proposition (4). It does not necessarily imply that after equating marginal cost with marginal revenue ($p^{c,\infty}$ and p^c), the

---

24

$$\theta(1, 1 - d_a^c) \geq \theta^\infty(\frac{R_{\phi_t^c=1}}{R^c}, d_a^{c,\infty}) \iff 1 - e^{-(1+e^{1-d_a^c})} \geq 1 - e^{-(\frac{R_{\phi_t^j=1}}{R^j} + e^{d_a^{c,\infty}})} \iff$$
$$(1 + e^{1-d_a^c}) \geq \frac{R_{\phi_t^j=1}}{R^j} + e^{d_a^{c,\infty}} \iff 1 - d_a^c \geq d_a^{c,\infty}$$

volume of email sent will decrease. This is because with the introduction of the open channel, senders benefit from complementarity so that the marginal cost of sending is cheaper. In all, to the extent that the complementarity effect does not crowd out the effect from the decrease in marginal revenue, the volume of censored-content mass mail sent will decrease. The associated dis-utility in turn will have to depend on the tradeoff between the preference of information over volume as detailed by the $I(\cdot)$ in the recipient's problem.

Now we make one simplifying assumption before we analyze the welfare effects.

Assumption 6 *The uncensored-content mail senders do not change the email volume they sent to the censored channel after the open channel is introduced. That is,* $\hat{n}_r^c = \hat{n}_r^{c,\infty}$.

Assumption 3 already implied that the uncensored-content mail senders find it profit-maximizing to use only the censored channel. Assumption 6 further asserts that when the open channel is absent, the censored channel is no less or more productive for such senders. Of course, in reality, it is possible to have a scenario that we nonetheless ruled out below. When the open channel is absent, there could be much censored-content mail in the censored channel (as we have shown in Proposition 5) that could interfere with other communications such as uncensored-content mail. Thus, with the introduction of the open channel, the censored channel is becoming more productive for uncensored-content mail senders, and it is possible that $\hat{n}_r^c \geq \hat{n}_r^{c,\infty}$. To be more precise, we rule this out because we want to have a benchmark to compare the welfare with and without the open channel. If one allows $\hat{n}_r^c \geq \hat{n}_r^{c,\infty}$ to happen, then there might be only welfare loss when the open channel simply because $\hat{n}_r^c$ becomes very high and the dis-utility associated with the portion of this that is wrongly filtered (at a rate of $\hat{\gamma}^c$) completely offset the utility associated with more (less) spam received for recipients who (do not) want it. We therefore think Assumption 6 as a benchmark is reasonable.

Proposition 6 *Under Propositions 1, 2, 4, 5, Assumptions 1, 2, 3, 4, and a sufficient condition, the welfare of the advertisers, censored-content mass-mail senders and all recipients will be unchanged or increased when there is an open channel.*

Proof. The welfare of the censored-content 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 in Appendix 6.3[25]:

$$U^r - \lim_{\gamma^o, \hat{\gamma}^o \longrightarrow \infty} U^r \tag{18}$$

Our interpretation is that when the filter strengths associated with the open channel are infinitely strong, it is as if there is no such channel for any practical use. ■

Essentially, we have proved that each user's utility has not decreased after the roll-out of the open channel because the utility associated with desired mail received does not decrease, and the dis-utilities associated with Type 1 and 2 errors do not increase.

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? 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 with the sides of buyers and sellers. 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)?[26] Currently, Gmail lists

[25] Technically, the limit is well defined only if the disguise levels converge to zero faster than the filter strengths to diverge to infinity (note: the bounds of a disguise level is zero and 1 in the limit) because some of the denominators could otherwise become zero's. But the disguise levels will indeed not tend to zero. Since the disguise cost is assumed to be convex, given a fixed advertising price, the senders cannot make enough money to justify the increasing disguise costs when the disguise levels tend to zero.

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

side-by-side some advertisements even for some 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?

- Will the open channel be flooded with email? Even if yes, the email is not disguised. Will a search function sufficiently offset the inconvenience of such flooding? Also, will spammers, who might no longer disguise their identities, weed out each other (and their associated volume) by establishing reputation? Lastly, will the Type-2 errors associated with spam sufficiently small (so flooding does not cause much inconvenience) as we have subsumed in the definition of spam?

- What are the other reasons to stop spammers from sending persuasive advertising to the censored channel? Will the undisguised mail in the open channel be sufficient threats to any persuasive mail in the censored channel? Will sellers utilizing the open channel be able to undercut the price of any goods or services sold in the censored channel since there is no need for them to incur disguise costs?

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 recipients, spammers, and advertisers. 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. 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.

In our mathematical model, we have shown that under certain conditions, 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 both undesirable mail received and desirable mail filtered out decreases. We especially have taken into account of how recipients trade off between information and volume by introducing the concept of deviations from information neutrality.

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 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 (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 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 the right to recipients to receiving 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 also 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 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 is conceivable that it would make email users at least weakly better off (no worse off) than the status quo. At the very least, the solution 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 Proposition 2

The sender's profit function is

$$\pi_m(n_m^o, n_m^c, d_m^c) = p^o n_m^o + \frac{p^c n_m^c}{d_m^c \gamma^c} - c_m(n_m^o, n_m^c, d_m^c), \tag{19}$$

the Lagrangian is:

$$\mathcal{L} = \pi(\cdot) - \lambda_1^c(d_m^c - 1) + \lambda_2^c(d_m^c - \frac{1}{\gamma^c}) + \mu^o n_m^o + \mu^c n_m^c \tag{20}$$

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

The complementary slackness conditions are:

$$\lambda_1^c(d_m^c - 1) = 0 \tag{21}$$

$$\lambda_2^c(d_m^c - \frac{1}{\gamma^c}) = 0 \tag{22}$$

$$\mu^o n_m^o = 0 \tag{23}$$

$$\mu^c n_m^c = 0 \tag{24}$$

FOCs:

$$\begin{aligned} p^o = \frac{\partial c_m}{\partial n_m^o} - \mu^o &= \delta n_m^c + n_m^o - \mu^o \\ \implies n_m^o &= p^o - \delta n_m^c + \mu^o \end{aligned} \tag{25}$$

$$\begin{aligned} \frac{p^c}{d_m^c \gamma^c} = \frac{\partial c_m}{\partial n_m^c} - \mu^c &= \delta n_m^o + n_m^c - \mu^c \\ \implies n_m^c &= \frac{p^c}{d_m^c \gamma^c} - \delta n_m^o + \mu^c \end{aligned} \tag{26}$$

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

$$1 = \frac{p^c n_m^c}{\gamma^c} + (\lambda_1^c - \lambda_2^c)(d_m^c)^2 \tag{27}$$

$$\implies d_m^c = (\frac{1 - \frac{p^c n_m^c}{\gamma^c}}{\lambda_1^c - \lambda_2^c})^{1/2}, \lambda_1^c \neq \lambda_2^c \tag{28}$$

Combining (25) and (26):

$$\begin{aligned} n_m^o &= p^o - \delta(\frac{p^c}{d_m^c \gamma^c} - \delta n_m^o + \mu^c) + \mu^o \\ &= \frac{1}{1-\delta^2}[p^o - \delta(\frac{p^c}{d_m^c \gamma^c} + \mu^c) + \mu^o] \end{aligned} \tag{29}$$

$$\begin{aligned} n_m^c &= \frac{p^c}{d_m^c \gamma^c} - \delta(p^o - \delta n_m^c + \mu^o) + \mu^c \\ &= \frac{1}{1-\delta^2}[\frac{p^c}{d_m^c \gamma^c} - \delta(p^o + \mu^o) + \mu^c] \end{aligned} \tag{30}$$

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 (29),

$$n_m^o = \frac{1}{1-\delta^2}[p^o - \delta(\frac{p^c}{d_m^c \gamma_m^c})] \tag{31}$$

From (30),

$$n_m^c = \frac{1}{1-\delta^2}[\frac{p^c}{d_m^c \gamma^c} - \delta p^o] \tag{32}$$

Subcase 1: $d_m^{*c} = 1 \implies \lambda_2^c = 0$
From (31),

$$n_m^{*o} = \frac{1}{1-\delta^2}[p^o - \delta(\frac{p^c}{\gamma_m^c})] \tag{33}$$

From (32),

$$n_m^{*c} = \frac{1}{1-\delta^2}[\frac{p^c}{\gamma^c} - \delta p^o] \tag{34}$$

From (27),

$$\lambda_1^c = \frac{\gamma^c - p^c n_m^{*c}}{\gamma^c} \tag{35}$$

From (35), $\lambda_1^c > 0 \iff$

$$\gamma^c > p^c n_m^{*c} \tag{36}$$

$$\gamma^c > \frac{p^c}{1-\delta^2}(\frac{p^c}{\gamma^c} - \delta p^o) \tag{37}$$

From (35), $\lambda_1^c = 0 \iff$

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

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

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 $p^o \leq \frac{p^c}{\delta \gamma^c} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$.[27]
Subcase 2: $d_m^{*c} = \frac{1}{\gamma^c} \implies \lambda_1^c = 0$
From (31),

$$n_m^{*o} = \frac{1}{1-\delta^2}[p^o - \delta p^c] \tag{38}$$

From (32),

$$n_m^{*c} = \frac{1}{1-\delta^2}[p^c - \delta p^o] \tag{39}$$

From (27),

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

$$1 = \frac{p^c n_m^{*c}}{\gamma^c} - \frac{\lambda_2^c}{(\gamma^c)^2} \tag{40}$$
$$\lambda_2^c = \gamma^c p^c n_m^{*c} - (\gamma^c)^2 \tag{41}$$

From (41), $\lambda_2^c > 0 \iff$

$$\gamma^c p^c n_m^{*c} - (\gamma^c)^2 > 0 \tag{42}$$
$$p^c n_m^{*c} > \gamma^c \tag{43}$$
$$\frac{p^c(p^c - \delta p^o)}{1-\delta^2} > \gamma^c \tag{44}$$

From (41), $\lambda_2^c = 0 \iff$

$$p^c n_m^{*c} = \gamma^c \tag{45}$$
$$\frac{p^c(p^c - \delta p^o)}{1-\delta^2} = \gamma^c \tag{46}$$

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 (27) and the premises for this subcase imply that:

$$\gamma^c = n_m^c p^c \tag{47}$$

Substitute (47) into (32) to get[28]:

$$d_m^{*c} = \frac{(p^c)^2}{(1-\delta^2)(\gamma^c)^2 + \delta p^o p^c \gamma^c} \tag{48}$$

28

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

By $d_m^{*c} \in (\frac{1}{\gamma^c}, 1)$, we must have

$$\frac{1}{\gamma^c} < \frac{(p^c)^2}{(1-\delta^2)(\gamma^c)^2 + \delta p^o p^c \gamma^c} < 1 \tag{49}$$

Or equivalently,

$$\frac{p^c}{\delta} - \frac{\gamma^c(1-\delta^2)}{\delta p^c} < p^o < \frac{p^c}{\delta\gamma^c} - \frac{\gamma^c(1-\delta^2)}{\delta p^c} \tag{50}$$

Note that the interval in which subcases 1 and 2 overlap is:

$$\frac{p^c}{\delta} - \frac{\gamma^c(1-\delta^2)}{\delta p^c} \leq p^o \leq \frac{p^c}{\delta\gamma^c} - \frac{\gamma^c(1-\delta^2)}{\delta p^c} \tag{51}$$

Note first that subcase 3 cannot coexist with either subcases 1 or 2 because the endpoints of the open interval of d_m^{*c} in subcase 3 coincides that of the other subcases. So the only possibility is that either of the following is true (but not both). (a) subcase 3 is true; (b) subcases 1 and (or) 2 are true. Suppose (a) is true, then (b) is not true. But this implies (50) and NOT (51) must be both true. This contradicts with (50) being subsumed in (51). Therefore, subcase 3 is eliminated.

Case 2: $n_m^{*o} > 0, n_m^{*c} = 0$

From (26),

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

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 (25),

$$n_m^o = p^o - \delta n_m^c + \mu^o \tag{52}$$

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

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

Case 4: $n_m^{*o} = n_m^{*c} = 0$.

From (25),

$$n_m^o = p^o - \delta n_m^c + \mu^o \tag{54}$$
$$0 = p^o + \mu^o \tag{55}$$

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

From (26),

$$n_m^c = \frac{p^c}{d_m^c \gamma^c} - \delta n_m^o + \mu^c \tag{56}$$
$$0 = \frac{p^c}{d_m^c \gamma^c} + \mu^c \tag{57}$$

Since the right hand side is non-negative, the only permissible values are $p^c = \mu^c = 0$.

$n_m^{*c} = 0$ implies that (28) gives:

$$d_m^{*c} = (\frac{1}{\lambda_1^c - \lambda_2^c})^{1/2} \tag{58}$$

Subcase 1: If $\lambda_1^c = 0$ and $\lambda_2^c = 0$, (27) implies a contradiction because:

$$\frac{p^c n_m^{*c}}{\gamma^c} = 1 \tag{59}$$
$$n_m^{*c} = \frac{\gamma^c}{p^c} \neq 0 \tag{60}$$

Subcase 2: If $\lambda_1^c = 0$ and $\lambda_2^c > 0$, it gives a contradiction of d_m^{*c} being negative by (58).

Subcase 3: If $\lambda_1^c > 0 \implies d_m^c = 1$ (see (21)).

Therefore, case 4 is admissible when

$$d_m^c = 1; p^o = p^c = 0 \tag{61}$$

Q.E.D.

6.2 Proof of Proposition 5

We will show that $n_m^{*c,\infty} \geq n_m^{*c}$ and $\frac{n_m^{*c,\infty}}{d_m^{*c,\infty}\gamma^c} \geq \frac{n_m^{*c}}{d_m^{*c}\gamma^c}$. And by homogeneity of senders and even distribution of mass mail, it is equivalent to proving $\frac{I(d_m^{*c,\infty})n_m^{*c,\infty}}{d_m^{*c,\infty}\gamma^c} \geq \frac{I(d_m^{*c})n_m^{*c}}{d_m^{*c}\gamma^c}$ to show that $\frac{I(d^{*c,\infty})n_r^{*c,\infty}}{d^{*c,\infty}\gamma^c} \geq \frac{I(d^{*c})n_r^{*c}}{d^{*c}\gamma^c}$ because $\frac{1}{R^c}\sum_m n_m^{c,\infty} \equiv n_r^{c,\infty}$ and $\frac{1}{R^c}\sum_m n_m^c \equiv n_m^c$.

Recall that when the open channel is present:

Case (a), $p^o \leq \frac{p^c}{\delta\gamma^c} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$:

$$n_m^{*c} = \frac{1}{1-\delta^2}(\frac{p^c}{\gamma^c} - \delta p^o) \tag{62}$$

$$\frac{n_m^{*c}}{d_m^{*c}\gamma^c} = \frac{n_m^{*c}}{\gamma^c} = \frac{1}{1-\delta^2}[\frac{p^c}{(\gamma^c)^2} - \frac{\delta p^o}{\gamma^c}] \tag{63}$$

$$\frac{I(d_m^{*c})n_m^{*c}}{d_m^{*c}\gamma^c} = \frac{I(1)n_m^{*c}}{\gamma^c} = \frac{1}{1-\delta^2}[\frac{p^c}{(\gamma^c)^2} - \frac{\delta p^o}{\gamma^c}] \tag{64}$$

Case (b), $p^o \geq \frac{p^c}{\delta} - \frac{\gamma^c(1-\delta^2)}{\delta p^c}$:

$$\frac{n_m^{*c}}{d_m^{*c}\gamma^c} = n_m^{*c} = \frac{1}{1-\delta^2}(p^c - \delta p^o) \tag{65}$$

$$\frac{I(d_m^{*c})n_m^{*c}}{d_m^{*c}\gamma^c} = I(\frac{1}{\gamma^c})n_m^{*c} = \frac{1}{1-\delta^2}[\frac{p^c}{(\gamma^c)^\beta} - \frac{\delta p^o}{(\gamma^c)^\beta}] \tag{66}$$

Again we use ∞ to denote variables when the open channel is absent. When the open channel is absent, it can be shown that we could simply substitute $\delta = 0$, and $p^o = 0$ (or equivalently $\gamma^o \to \infty$) into the above equations and the admissible ranges (but first multiply both sides with δ for the ranges before the substitutions so the ranges will not be undefined):

Case (a), $p^{c,\infty} \leq \gamma^c$:

$$n_m^{*c,\infty} = \frac{p^{c,\infty}}{\gamma^c} \tag{67}$$

$$\frac{n_m^{*c,\infty}}{d_m^{*c,\infty}\gamma^c} = \frac{n_m^{c,\infty}}{\gamma^c} = \frac{p^{c,\infty}}{(\gamma^c)^2} \tag{68}$$

$$\frac{I(d_m^{*c,\infty})n_m^{*c,\infty}}{d_m^{*c,\infty}\gamma^c} = \frac{I(1)n_m^{c,\infty}}{\gamma^c} = \frac{p^{c,\infty}}{(\gamma^c)^2} \tag{69}$$

Case (b), $p^{c,\infty} \geq (\gamma^c)^{1/2}$:

$$\frac{n_m^{*c,\infty}}{d_m^{*c,\infty}\gamma^c} = n_m^{*c,\infty} = p^{c,\infty} \tag{70}$$

$$\frac{I(d_m^{*c,\infty})n_m^{*c,\infty}}{d_m^{*c,\infty}\gamma^c} = I(\frac{1}{\gamma^c})n_m^{*c,\infty} = \frac{p^{c,\infty}}{(\gamma^c)^\beta} \tag{71}$$

Without further specifying more functional forms of θ^j and k in the advertiser's problem, it is not possible to decide which case the parameters will fall into under the scenarios of whether the open channel is present or absent, it may not be appropriate to compare the same case across scenarios. For example, it may not be appropriate to compare the case (a) with the open channel with the case (a) without the open channel. But we know that the case (b) variables at issues are always weakly larger with and without the open channel when $\beta \in [0,1]$. It is then sufficient to show that the variables at issue in case (a) without the open channel is always weakly larger than the counterpart variables in case (b) with the open channel for $\beta \in [0,1]$. Essentially, we will show a sufficient condition so that:

$$n^{*c,\infty}_{m,case\ a} \geq n^{*c}_{m,case\ b} \tag{72}$$

$$\frac{n^{*c,\infty}_m}{d^{*c,\infty}_m \gamma^c}_{case\ a} \geq \frac{n^{*c}_m}{d^{*c}_m \gamma^c}_{case\ b} \tag{73}$$

$$\frac{I(d^{*c,\infty}_m) n^{*c,\infty}_m}{d^{*c,\infty}_m \gamma^c}_{case\ a} \geq \frac{I(d^{*c}_m) n^{*c}_m}{d^{*c}_m \gamma^c}_{case\ b} \tag{74}$$

In fact, the second inequality implies the third because $I(d^{*c,\infty}_m)_{case\ a} = 1 \geq I(d^{*c}_m)_{case\ b} = \frac{1}{(\gamma^c)^\beta}$.

Equivalently, the second inequality can be rewritten as:

$$\frac{p^{c,\infty}}{(\gamma^c)^2} \geq \frac{1}{1-\delta^2}(p^c - \delta p^o) \tag{75}$$

And the first inequality can be rewritten as:

$$\frac{p^{c,\infty}}{\gamma^c} \geq \frac{1}{1-\delta^2}(p^c - \delta p^o) \tag{76}$$

Note that the second inequality also implies the first inequality because (75) $\implies$ (76).

Applying the first-order condition of the advertisers' problem, we have $p^o = \theta^o(s-k'); p^{c,\infty} = \theta^{o,\infty}(s-k^{\infty\prime})$. So $\frac{1}{1-\delta^2}(p^c - \delta p^o) = \frac{1}{1-\delta^2}(\theta^c - \delta\theta^o)(s-k')$ and (75) becomes:

$$\frac{\theta^{o,\infty}(s-k^{\infty\prime})}{(\gamma^c)^2} \geq \frac{1}{1-\delta^2}(\theta^c - \delta\theta^o)(s-k'), \tag{77}$$

where $\frac{R_{\phi^o_t=1}}{R^o} = 1$ by Proposition 1. This is the first part of the sufficient condition for this proposition.

For $\beta > 1$, we have to additionally show the second part of the sufficient condition:

$$\frac{I(d^{*c,\infty}_m) n^{*c,\infty}_m}{d^{*c,\infty}_m \gamma^c}_{case\ a} \wedge \frac{I(d^{*c,\infty}_m) n^{*c,\infty}_m}{d^{*c,\infty}_m \gamma^c}_{case\ b} \geq \frac{I(d^{*c}_m) n^{*c}_m}{d^{*c}_m \gamma^c}_{case\ a} \wedge \frac{I(d^{*c}_m) n^{*c}_m}{d^{*c}_m \gamma^c}_{case\ b}, \tag{78}$$

or equivalently:

$$\frac{p^{c,\infty}}{(\gamma^c)^2} \wedge \frac{p^{c,\infty}}{(\gamma^c)^\beta} \geq \frac{1}{1-\delta^2}[\frac{p^c}{(\gamma^c)^2} - \frac{\delta p^o}{\gamma^c}] \wedge \frac{1}{1-\delta^2}[\frac{p^c}{(\gamma^c)^\beta} - \frac{\delta p^o}{(\gamma^c)^\beta}] \qquad (79)$$

Similarly, one could further substitute in the first-order condition of the advertisers' problem.

Q.E.D.

6.3 Proof of Proposition 6

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

$$\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} U^r(u^r_{\text{desired received}}, u^r_{\text{Type 1 errors}}, u^r_{\text{Type 2 errors}}) \qquad (80)$$

$$= U^r(\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{desired received}}, \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 1 errors}}, \lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 2 errors}}) \qquad (81)$$

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

$$\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{desired received}} = (1-\epsilon)\frac{\kappa^c_r \hat{n}^{c,\infty}_r}{\hat{\gamma}^c} + \frac{\phi^c_t \kappa^c_r I(d^{c,\infty}) n^{c,\infty}_r}{d^{c,\infty}\gamma^c} \qquad (82)$$

$$\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 1 errors}} = \epsilon\frac{\kappa^c_r \hat{n}^{c,\infty}_r}{\hat{\gamma}^c} + \frac{(1-\phi^c_t)\kappa^c_r I(d^{c,\infty}) n^{c,\infty}_r}{d^{c,\infty}\gamma^c} \qquad (83)$$

$$\lim_{\gamma^o,\hat{\gamma}^o \longrightarrow \infty} u^r_{\text{Type 2 errors}} \qquad (84)$$

$$= (1-\epsilon)\kappa^c_r \hat{n}^{c,\infty}_r (1 - \frac{1}{\hat{\gamma}^c}) + \phi^c_t \kappa^c_r \bar{I}(d^{c,\infty}) n^{c,\infty}_r (1 - \frac{1}{d^{c,\infty}\gamma^c}) \qquad (85)$$

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 prove each of these inequalities below.

(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:

$$\sum_{j\in\{o,c\}} (1-\epsilon)\kappa^j_r \frac{\hat{n}^j_r}{\hat{\gamma}^j} + \sum_{j\in\{o,c\}} \phi^j_t \kappa^j_r \frac{I(d^j) n^j_r}{d^j \gamma^j}$$

$$\geq (1-\epsilon)\frac{\kappa^c_r \hat{n}^{c,\infty}_r}{\hat{\gamma}^c} + \frac{\phi^c_t \kappa^c_r I(d^{c,\infty}) n^{c,\infty}_r}{d^{c,\infty}\gamma^c} \qquad (86)$$

Assumption 3 implies that $(1-\epsilon)\kappa_r^o \frac{\hat{n}_r^o}{\hat{\gamma}^o} = 0$. Assumption 6 implies that $(1-\epsilon)\kappa_r^c \frac{\hat{n}_r^c}{\hat{\gamma}^c} = (1-\epsilon)\frac{\kappa_r^c \hat{n}_r^{c,\infty}}{\hat{\gamma}^c}$. (86) becomes:

$$\frac{\phi_t^c I(d^o)\kappa_r^o n_r^o}{d^o \gamma^o} \geq \frac{\phi_t^c \kappa_r^c I(d^{c,\infty}) n_r^{c,\infty}}{d^{c,\infty}\gamma^c} \tag{87}$$

The above is true for $\phi_t^o = 0$ because both sides are zero. For $\phi_t^o = 1$, $\kappa_r^o = \kappa_r^c = 1$ by Proposition 1. Note that $d^o = 1$, $\gamma_s^o = 1$, and $I(1) = 1$. (87) becomes:

$$n_r^o \geq n_r^{c,\infty} \tag{88}$$

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 $\frac{p^{c,\infty}}{d_m^{c,\infty}\gamma^c}$ and the marginal cost is $n_m^{c,\infty}$ (set $\delta = 0$ in $\frac{\partial c_m}{\partial n_m^{c,\infty}} = \delta n_m^{o,\infty} + n_m^{c,\infty}$). Profit maximization implies that each censored-content mass-mail sender equates marginal revenue with marginal cost:

$$p^o = \delta n_m^c + n_m^o \tag{89}$$

and

$$\frac{p^{c,\infty}}{d_m^{c,\infty}\gamma^c} = n_m^{c,\infty} \tag{90}$$

By Proposition 4, $p^o \geq p^{c,\infty}$. Together with $d^{c,\infty}\gamma^c \geq 1$, we have

$$p^o \geq \frac{p^{c,\infty}}{d_m^{c,\infty}\gamma^c} \tag{91}$$

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

$$\delta n_m^c + n_m^o \geq n_m^{c,\infty} \iff \tag{92}$$

$$n_m^o \geq n_m^{c,\infty} \text{ because } \delta < 0 \iff \tag{93}$$

$$\sum_m n_m^o \geq \sum_m n_m^{c,\infty} \text{ by homogeneity of senders } \iff \tag{94}$$

$$\frac{\sum_m n_m^o}{R^o} \geq \frac{\sum_m n_m^{c,\infty}}{R^{c,\infty}} \text{ since } R^{c,\infty} \geq R^o \text{ by Proposition 1 } \iff \tag{95}$$

$$n_r^o \geq n_r^{c,\infty} \text{ because by definition } \frac{\sum_m n_m^j}{R^j} = n_r^j \tag{96}$$

(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:

$$\sum_{j\in\{o,c\}} \epsilon\kappa_r^j \frac{\hat{n}_r^j}{\hat{\gamma}^j} + \sum_{j\in\{o,c\}} (1-\phi_t^j)\kappa_r^j \frac{I(d^j)n_r^j}{d^j\gamma^j}$$
$$\leq \epsilon\frac{\kappa_r^c \hat{n}_r^{c,\infty}}{\hat{\gamma}^c} + \frac{(1-\phi_t^c)\kappa_r^c I(d^{c,\infty})n_r^{c,\infty}}{d^{c,\infty}\gamma^c} \tag{97}$$

Assumption 3 implies that $\epsilon\frac{\kappa_r^o \hat{n}_r^o}{\hat{\gamma}^o} = 0$. Assumption 6 implies that $\epsilon\frac{\kappa_r^c \hat{n}_r^c}{\hat{\gamma}^c} = \epsilon\frac{\kappa_r^c \hat{n}_r^{c,\infty}}{\hat{\gamma}^c}$. (97) becomes:

$$\sum_{j\in\{o,c\}} (1-\phi_t^j)\kappa_r^j \frac{I(d^j)n_r^j}{d^j\gamma^j} \leq \frac{(1-\phi_t^c)\kappa_r^c I(d^{c,\infty})n_r^{c,\infty}}{d^{c,\infty}\gamma^c} \tag{98}$$

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 sufficient to prove for all low type recipients that:

$$\frac{I(d^c)n_r^c}{d^c} \leq \frac{I(d^{c,\infty})n_r^{c,\infty}}{d^{c,\infty}}, \tag{99}$$

which has already been proved under a sufficient condition in Appendix 6.2.

(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:

$$\sum_{j\in\{o,c\}} (1-\epsilon)\kappa_r^j \hat{n}_r^j (1-\frac{1}{\hat{\gamma}^j}) + \sum_{j\in\{o,c\}} \phi_t^j \kappa_r^j \bar{I}(d^j) n_r^j (1-\frac{1}{d^j\gamma^j})$$
$$\leq (1-\epsilon)\kappa_r^c \hat{n}_r^{c,\infty}(1-\frac{1}{\hat{\gamma}^c}) + \phi_t^c \kappa_r^c \bar{I}(d^{c,\infty}) n_r^{c,\infty}(1-\frac{1}{d^{c,\infty}\gamma^c}) \tag{100}$$

Assumption 3 implies that$(1-\epsilon)\kappa_r^o \hat{n}_r^o(1-\frac{1}{\hat{\gamma}^o}) = 0$. Assumption 6 implies that $(1-\epsilon)\kappa_r^c \hat{n}_r^c(1-\frac{1}{\hat{\gamma}^c}) = (1-\epsilon)\kappa_r^c \hat{n}_r^{c,\infty}(1-\frac{1}{\hat{\gamma}^c})$. (100) becomes:

$$\sum_{j\in\{o,c\}} \phi_t^j \kappa_r^j \bar{I}(d^j) n_r^j (1-\frac{1}{d^j\gamma^j}) \leq \phi_t^c \kappa_r^c \bar{I}(d^{c,\infty}) n_r^{c,\infty}(1-\frac{1}{d^{c,\infty}\gamma^c}) \tag{101}$$

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 2, (101) becomes:

$$\sum_{j=o,c} \bar{I}(d_s^j) n_r^j (1-\frac{1}{d^j\gamma^j}) \leq \bar{I}(d^{c,\infty}) n_r^{c,\infty}(1-\frac{1}{d^{c,\infty}\gamma^c}) \tag{102}$$

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

$$\bar{I}(d^c)n_r^c(1 - \frac{1}{d^c\gamma^c}) \le \bar{I}(d^{c,\infty})n_r^{c,\infty}(1 - \frac{1}{d^{c,\infty}\gamma^c}) \tag{103}$$

By Appendix 6.2, we already knew that under a sufficient condition, $n_r^c \le n_r^{c,\infty}$, and since $\bar{I}(\cdot)$ is an increasing function, we just need to show $d^c \le d^{c,\infty}$ to prove (103). Since from Appendix 6.2, we already knew that d^c and $d^{c,\infty}$ are either 1 in case (a) or $\frac{1}{\gamma^c}$ in case (b) depending on the parameters that determined which cases will happen with and without the open channel. Without the actual values of the parameters or more specific functional forms, it is possible to tell whether we should compare case (a) with the open channel with case (a) or case (b) without the open channel, etc. We therefore will state a sufficient condition for $d^c \le d^{c,\infty}$ to hold. This sufficient condition is that the case (a) in the with-open-channel scenario implies the case (a) in the without-open-channel scenario, and that the case (b) in the without-open-channel scenario implies the case (b) in the with-open-channel scenario.

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`.

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.