Spam - solutions and their problems

B. Curtis Eaton[*] Ian MacDonald[†] Laura Meriluoto[‡][§]

July 28, 2008

Abstract

We analyze the success of filtering as a solution to the spam problem when used alone or concurrently with sender and/or receiver pricing. We find that filters alone may exacerbate the spam problem if the spammer attempts to evade them by sending multiple variants of the message to each consumer. Sender and receiver prices can effectively reduce or eliminating spam, either on their own or when used together with filtering. Finally, we discuss the implications for social welfare of using the different spam controls.

Keywords: spam, filtering, email, receiver pricing, sender pricing

JEL classification numbers: L96, L10

[*] University of Calgary; eaton@ucalgary.ca.

[†] Lincoln University; macdonai@lincoln.ac.nz.

[‡] University of Canterbury; laura.meriluoto@canterbury.ac.nz.

[§] We would like to thank the conference participants in the 2007 New Zealand Association of Economists Conference, the 2007 E.A.R.I.E. Conference and the 2008 Canadian Economic Association Meetings for their helpful comments.

1 Introduction

Unsolicited commercial email or 'spam' is an increasingly significant problem for the email users and their network providers. It is estimated that spam currently accounts for as much as of 90% of all email traffic (The Economist, 2007), up from only 50% in 2003 and 7% in 2001 (US Public Law, 2003). This huge increase in email volume has imposed costs on internet service providers (ISPs) associated with wasteful consumption of bandwidth, increased demand on mail servers and a corresponding decrease in processor performance and has necessitated investment in increased infrastructure that would not otherwise be required. The users of the email network are also adversely affected by spam and incur direct costs associated with the processing of spam, indirect costs resulting from decreased speed and reliability of email systems[1], and psychological costs associated with the receipt of offensive messages or an overwhelming number of emails.

Spam exists because, from a business perspective, it works. Because spammers are incapable of identifying who their potential customers are ex ante, few people who are contacted by spammers are interested in the products on offer. While this means that expected benefit of a spam message is likely to small, spammers can still be profitable because the marginal cost of sending an email message is extremely small. The process of sending millions of untargeted messages can be profitable for spammers with response rates as low as 0.01% (The Economist, 2007).

Many countries, including the USA, Canada, New Zealand, India, and the countries of the European Union, have taken a regulatory approach to controlling spam. The US CAN-SPAM legislation passed in 2004, for example, imposes hefty fines on individuals or companies within the USA that send unwanted commercial email (US Public Law, 2003). However, even though there have been some convictions under legislation of this sort, it is unlikely to provide widespread relief from spam for two reasons. First, successful enforcement requires that the sender and receiver of a message be in the same jurisdiction which, in turn, requires that spammers must not be able to relocate to countries with no anti-spam laws. Second, successful enforcement requires that spammers cannot hide their true identity through spoofing or the practice of using viruses to illegally hijack consumers' computers turning them into 'spam zombies' (Griffiths, 2006).

[1] For example, tens of thousands of New Zealanders experienced 24 hour delays in receiving emails when their ISP was bombarded by spam messages that were not caught by its filters (Chug, 2006).

A number of technological defenses designed to filter or block unwanted messages from consumers' inboxes are available but these too have proven to be ineffective at eliminating spam. Blacklisting blocks messages sent by specific senders who have been identified as undesirable. Exclusive whitelisting blocks all messages except those coming from specific senders identified as acceptable. Nonexclusive whitelisting ensures that all messages from identified senders are allowed through any filters. Content based filtering blocks messages based on the message's subject matter and/or subject heading. The effectiveness of all three approaches in removing unwanted emails is constrained by the need to avoid removing messages that are wanted. In other words, there is the need to find a balance between allowing spam messages through (false negatives) and avoiding the capture of non-spam messages (false positives). With blacklisting there is essentially no chance of false positives but the process is completely ineffective if spammers can easily hide or quickly change their identities. Exclusive whitelisting is very effective at blocking spam but eliminates the scope for email to be used as a widespread means of communication between people who don't know each other. Nonexclusive whitelisting does little to reduce spam but does reduce the problem of false negatives associated with filters. With filters there is some scope for adjusting their stringency, which in turn affects the likelihood of false negatives and false positives, but spammers can attempt to evade filters by hiding the true subject or content of a message either by adding characters to disguise certain keywords or sending messages as images rather than text. Spammers may also send a large number of variant messages to each consumer in the hope that at least one of them will evade capture by the filters.

It is important to recognize that even if spam messages are blocked from consumers' inboxes ISPs need to process all messages sent by spammers. This means that for filters to be truly effective they must not only reduce the number of messages that arrive in consumers' inboxes but also reduce the total number of messages that are sent by spammers. We show in this paper that if spammers can increase the likelihood of evading filters by sending multiple variants of a message to each consumer, it is entirely possible that filtering will actually exacerbate the problem of spam on both fronts.

Economic defences against spam discussed in the literature include sender pricing (see for example Arrison (2004), Dai and Li (2004), Khong (2004) and Kraut et. al. (2005)) and attention bonds (see for example Fahlman (2002), Loder, Van Alstyne and Wash (2006) and Van Alstyne (2007)) but these methods have yet to be used in practice. The literature suggests that a sender price in the order of fractions of cents per message could eliminate spam by increasing spammers' per message costs above their expected per message revenue. Likewise an attention bond that grants the recipient of a message a right to set a fee for their attention, payable if the receiver decides that the sender was wasting her time, might also be effective.

We construct a model of a monopolist spammer and a single ISP provider to examine the impact of filters and prices on the spammer's choice of i) the number of variant messages sent to each targeted consumer and ii) the number of consumers to target. We show that receiver pricing could reduce or eradicate spam by reducing the number of consumers who will read spam messages therefore reducing the spammer's expected marginal benefit of sending spam. Similarly sender pricing could reduce or eradicate spam by increasing the spammer's cost per message. We show that there is a real possibility that filters used on their own will lead to a manyfold increase in the total volume of spam, such that the expected number of spam messages that evade filters and end up in targets' inboxes could actually increase compared to a situation when filtering is not used at all.

Our goal in this paper is not to model the game played between those who's objective it is to design effective filters and spammers who's objective is to evade them. Clearly this is an involved and dynamic process that warrants further analysis. Instead we assume that the spammer takes the effectiveness of filters as an exogenous parameter and therefore in order to increase the likelihood of evading filters, he must send multiple variants of a message to each target. Specifically our filter blocks each and every message with probability q and so if n messages are sent to a particular target, the spammer's likelihood of getting at least one message through the filter into the target's inbox is $(1 - q^n)$. When the expected benefit of making contact with a target is large compared to the cost of sending messages filtering can lead to an increase in the total volume of spam. The problem that filters may cause, however, can be mitigated through the use of email pricing because receiver pricing reduces the spammer's expected benefit of making contact with a target and sender pricing increases the cost of sending messages.

We also analyze the comparative statics of the number of variant messages sent to each targeted consumer and the number of consumers targeted with respect to changes in the magnitude of the receiver and sender prices and the effectiveness of the filter. We find that the magnitude of the spam-eliminating receiver and sender prices are inversely related to the effectiveness of the filter suggesting that filters and prices complement each other in the fight against spam.

Most of the existing literature in the area is concerned with the welfare effects of spam. Hermalin and Katz (2004) examine efficient pricing of email generally but do not focus on the problem of spam. Shiman (1996, 2006) and Ayres and Funk (2003) analyze conditions under which spam is likely to reduce social welfare. Loder, Van Alstyne and Wash (2006) analyze the welfare effects of three competing economic responses to unsolicited email in a model in which the utility of some messages does not exceed their costs: a flat tax is used to internalize the external effect of a message to the receiver, attention bonds are used to directly compensate the receivers for their costs of reading messages, and a perfect filter is used to capture all unwanted messages. The filter is assumed to be perfect in that it is able to capture any and all messages with value less than the processing cost to the recipient but allows all other messages to escape. Filtering reduces the number of spam messages that are sent because the spammer's expected value of a message is reduced as a consequence of the reduction in the proportion of spam messages that are received by consumers[2].

Although the primary focus of this paper is on the effect of filtering and email pricing on the volume of spam, we do provide a brief analysis of the possible welfare effects of these controls. If a sender price can be levied exclusively on spam and if spam is welfare reducing we find that welfare unambiguously increases with an increase in the sender price. A receiver price that is similarly levied exclusively on spam does not have the same unambiguous welfare effect. If sender and receiver prices cannot be directly targeted to spam, however, they will impact on all network activity and this means that any welfare gains from controlling spam, if indeed they exist, could be offset by welfare losses associated with less communication between non-spammers.

The paper is structured as follows. Section 2 introduces the formal model consisting of a single ISP and a monopolist spammer who chooses the size of his mailing list in stage 1 and the number of message variants to send to each consumer in stage 2. Using this framework, we determine the impact of filtering and pricing on the number of messages sent to each

[2] This is based on the assumption made by Loder et. al. that the spammer receives a benefit of getting messages through to a consumer even if that consumer is not interested in its product.

target and the number of messages that each target receives in her inbox in Section 3. In Section 4 we examine how filtering and pricing affect the size of the spammer's mailing list and, consequently, the total volume of spam that he sends. In Section 5 we examine the possible welfare effects of sender and receiver prices. Section 6 concludes.

2 The model

This section presents a model of monopoly spammer and describes its profit maximizing choice of the number of consumers to contact and the number of message variants to send to each contact. We determine how these choices, as well as the expected number of messages arriving in a target's inbox, are affected by filtering, receiver pays pricing and sender pays pricing. We allow the spammer limited scope for avoiding these anti-spam measures. Specifically, we assume that the spammer can only send multiple variants of a message to each target in an attempt to evade filtering. Moreover, because we are focusing here only on spammer behavior and are not concerned with modelling ISP decisions per se, we treat the ISP as if it were a single autonomous entity that services all participants in the email network.

In our model the spammer is interested in selling his product to consumers and, in order to do so, must make contact with a consumer who is interested in purchasing the product. In order for such a contact to be made two things must occur. First, the spammer must place an email message in the consumer's inbox by both utilizing their address and eluding any spam filters that are in place. Second, the interested consumer must read the message[3]. Importantly, we assume that consumers cannot be identified by their tastes for spam and so the spammer cannot target his messages. Instead the spammer must contact consumers at random and this indiscriminate sending of messages means that for every message that finds its way into an interested buyer's inbox, many more are likely to be filtered or received by uninterested consumers.

Each spam message that is sent costs the spammer c^{spam} to process and send[4]. This per message cost for the spammer is certainly small and likely to be very close to zero. Each spam message sent costs the ISP c^U to transmit. Each message that arrives in a consumer's inbox

[3] We make a distinction between receiving a message in one's inbox and reading a message. The spammer receives utility of its message for only those consumers who read the message because they are the only ones who get the spammer's message.

[4] In reality many of the spammers costs (such as access/bandwidth, labor, hardware, development of ways to avoid filtering, etc.) will be lumpy. For simplicity we model them as a constant per message marginal cost.

costs that consumer c^R to process, where processing involves either opening, filing and/or deleting the message. Thus, c^R is sunk at the time the receiver decides whether or not to open and read the message as it must be incurred regardless of the decision.

The spammer pays a per message sender price $p^S \geq -c^{spam}$ to the ISP and the receiver of a message pays a per message receiver price of $p^R \geq 0$ to the ISP upon opening a message. The restriction on the sender price ensures that there is no incentive to manufacture and send phoney messages as, in effect, a commercial activity. We place two restrictions on p^R in order to rule out receiver prices that cannot influence the behavior of receivers in a useful way. First, because opening a message is not tantamount to reading the message, negative receiver pays pricing cannot be used to induce consumers to read messages that they would not otherwise choose to read and so are ruled out here. Second, we assume that consumers are only required to pay the receiver price for those messages in their inbox that they choose to open.

Although the spammer does not know the preferences of any particular consumer, he does have complete information about the benefit interested consumers receive from his message and about the magnitude of the receiver price. This means that the spammer can determine *ex ante* whether or not his messages will be read by those consumers who are interested. The spammer makes a profit of π associated with making a sale from each interested customer that reads his spam message.

If consumer i reads a spam message she receives utility ρ_i^{spam} drawn from a smooth and continuous distribution $f(\rho^{spam})$ with support in $[\rho_{min}^{spam} < 0, \rho_{max}^{spam} > 0]$. We denote the survival function as $(1 - F(\rho^{spam}))$. We assume that the message's heading and sender information contain enough information about a message for the consumer to infer its value[5]. Positive values of ρ_i^{spam} are associated with gaining valuable product information and reduced search costs so there is no value to a consumer of reading multiple spam messages. We assume that all consumers for whom $\rho_i^{spam} > 0$ will purchase one unit of the spammer's product if and only if they receive and read a spam message. Because c^R is sunk, consumer i will read a spam message and therefore purchase the spammer's product if $\rho_i^{spam} \geq p^R$. The proportion of consumers who will read a spam message if they receive at least one is $(1 - F(p^R))$, is

[5] Even if this assumption does not hold for all types of email messages, we believe that it is relatively straightforward to identify spam messages prior to opening and, given that they know a message is spam, most people have a good sense of its likely value. Even those spam messages that are spoofed typically have a mismatch in heading content between what the receiver would expect from the sender thus easing the identification of spam.

decreasing in p^R, and is equal to zero when $p^R = \rho^{spam}_{max}$.

Filtering technologies employed by the ISP block messages that are from particular origins or that contain certain words or phrases in the subject line or body of the message. The spammer does not know the exact filtering technologies employed by the ISP but can try to evade them by avoiding words or phrases that are likely to be caught and/or by sending a number of variants of the message, perhaps from different origins. We capture the essence of filtering by assuming that any message sent by a spammer has a probability q of being filtered and $(1 - q)$ of getting through to a consumer's inbox. By sending multiple variants of a message to a consumer, the spammer increases the likelihood of getting at least one message in the receiver's inbox. With n messages sent to a consumer, the probability of at least one message getting through to her inbox is $(1 - q^n)$. Each message the ISP is successful in filtering saves a consumer c^R but still costs the spammer and the ISP c^{spam} and c^U, respectively, to process.

The spammer has a two-stage problem. In stage 1, the spammer generates a mailing list using a process with increasing marginal cost of finding a unique name. Denote the size of the mailing list by M. In stage 2, the spammer chooses how many messages to send to each consumer on his mailing list or to each of his targets. Denote the number of messages sent by the spammer to each of the consumers on his list by n. Total volume of spam is simply nM. We use backward induction to solve the spammer's problem in the next two subsections and to determine how the spammer's choices are affected by filtering alone and together with receiver and sender pricing.

3 Stage 2 - Profit-maximizing number of messages per target

We introduce the stage 2 problem in continuous form even though it is not defined in the absence of filtering $(q = 0)$ and does not perform well when the optimal number of messages is less than one. We do this because the continuous model allows for the derivation of interesting closed-form comparative static results. However, because we believe that the discrete form of the model better represents the reality of the problem, particularly when the number of messages sent to each consumer is likely to be small, we also provide a discrete representation of the spammer's per target message choice in the appendix. Due to being well-defined at

$q = 0$, this discrete version of the problem has a more intuitive graphical interpretation than the continuous version. We illustrate the differences between the continuous problem and the discrete problem in Figures 2a and 2b.

The number of messages per target sent by the spammer is found by maximizing expected profit per consumer with respect to n:

$$\max_{\{n\}} \Pi = (1 - q^n)(1 - F(p^R))\pi - (c^{spam} + p^S)n. \tag{1}$$

The profit-maximizing number of messages per target is

$$n^* = \begin{cases} \frac{ln\left(-\frac{A}{ln(q)}\right)}{ln(q)} & \text{if } A \leq -\ln(q) \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where A is the ratio of the spammer's marginal cost and expected marginal revenue:

$$A = \frac{c^{spam} + p^S}{(1 - F(p^R))\pi}. \tag{3}$$

The expected number of messages received by each targeted consumer is

$$n^{inbox} = \begin{cases} \frac{(1-q)ln\left(-\frac{A}{ln(q)}\right)}{ln(q)} & \text{if } A \leq -\ln(q) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The maximized profit equals

$$\Pi(n^*) = \begin{cases} (1 - q^{\frac{ln\left(-\frac{A}{ln(q)}\right)}{ln(q)}})(1 - F(p^R))\pi - \frac{(c^{spam}+p^S)ln\left(-\frac{A}{ln(q)}\right)}{ln(q)} & \text{if } A \leq -\ln(q) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Notice for future reference that $\frac{\partial \Pi(n^*)}{\partial q} < 0$, $\frac{\partial \Pi(n^*)}{\partial p^S} < 0$ and that $\frac{\partial \Pi(n^*)}{\partial p^R} < 0$. Intuitively, filtering reduces the spammer's profit by reducing the likelihood of a message getting through to a consumer, sender price reduces the spammers profit directly and the receiver price reduces the spammers profit by reducing the response rate of the messages that end up in consumers' inboxes.

3.1 Spam-eliminating p^R and p^S

While it is not necessarily socially optimal to use prices that eliminate all spam, it is still useful to define the prices that would achieve this outcome.

The spam eliminating receiver price and sender price are the values of p^R and p^S, respectively, that set (2) equal to zero. The spam-eliminating receiver price is

$$p^R_{spam} = F^{-1}\left(1 + \frac{c^s + p^S}{ln(q)\pi}\right). \tag{6}$$

Notice that $p^R_{spam} \leq \rho^{spam}_{max}$ because spam becomes unprofitable with low positive response rates. p^R_{spam} is increasing in π and decreasing in q, c^s and p^S.

The spam-eliminating sender price is

$$p^S_{spam} = -(1 - F(p^R))\pi ln(q) - c^S. \tag{7}$$

p^S_{spam} is increasing in π and decreasing in q, c^s and p^R.

3.2 Comparative statics

In this subsection, we investigate how prices and the filter affect the number of messages sent when their levels are below the spam-eliminating levels. We also investigate the complementarity between these tools.

The profit-maximizing number of messages per target varies with p^R, p^S and q in the following ways:

$$\frac{\partial n^*}{\partial p^R} = \frac{\partial n^*}{\partial A}\frac{\partial A}{\partial p^R} = \frac{f(p^R)}{(1 - F(p^R))ln(q)} \leq 0, \tag{8}$$

$$\frac{\partial n^*}{\partial p^S} = \frac{\partial n^*}{\partial A}\frac{\partial A}{\partial p^S} = \frac{1}{(c^{spam} + p^S)\ln(q)} \leq 0 \tag{9}$$

and

$$\frac{\partial n^*}{\partial q} = -\frac{1 + \ln\left(-\frac{A}{\ln(q)}\right)}{\ln(q)^2 q}. \tag{10}$$

It is clear from (8) and (9) that receiver and sender pricing both unambiguously deter spam but that the impact of the effectiveness of the filter on the number of messages per target in (10) is ambiguous. To gain further insights into the interplay of q and prices in deterring spam, define $\hat{q}$ as the switching value of q that sets $\frac{\partial n^*}{\partial q} = 0$ for a given A:

$$\hat{q} \equiv e^{-Ae}. \tag{11}$$

Any increase in q will lead to an increase in the number of messages per target if $q < \hat{q}$ and a decrease in the number of messages if $q > \hat{q}$. That is not to say, however, that the number of messages per target will be lower for all $q > \hat{q}$ as compared to a situation with little or no filtering. Also recall that spam is costly to send and process regardless of whether it enters into consumers' inboxes and so the main conclusion to draw from the above analysis is that using filters on their own to control spam will not be prudent unless q is certain to be very large.

It is also evident, particularly in (10) and in (11), that the impact of filtering on both the number of messages sent per target and the number of messages that enter into consumers' inboxes depend on the level of sender and receiver prices. Differentiating $\hat{q}$ with respect to p^S yields

$$\frac{\partial \hat{q}}{\partial p^S} = \frac{\partial \hat{q}}{\partial A}\frac{\partial A}{\partial p^S} = -\frac{e^{1-Ae}}{(1-F(p^R))\pi} < 0. \tag{12}$$

This derivative shows that the critical effectiveness value is decreasing in the sender price and so filters are more likely to be an appropriate defense against spam if used in conjunction with a sender price than if used on their own. A similar result holds for p^R:

$$\frac{\partial \hat{q}}{\partial p^R} = \frac{\partial \hat{q}}{\partial A}\frac{\partial A}{\partial p^R} = -\frac{f(p^R)Ae^{1-Ae}}{(1-F(p^R))} < 0. \tag{13}$$

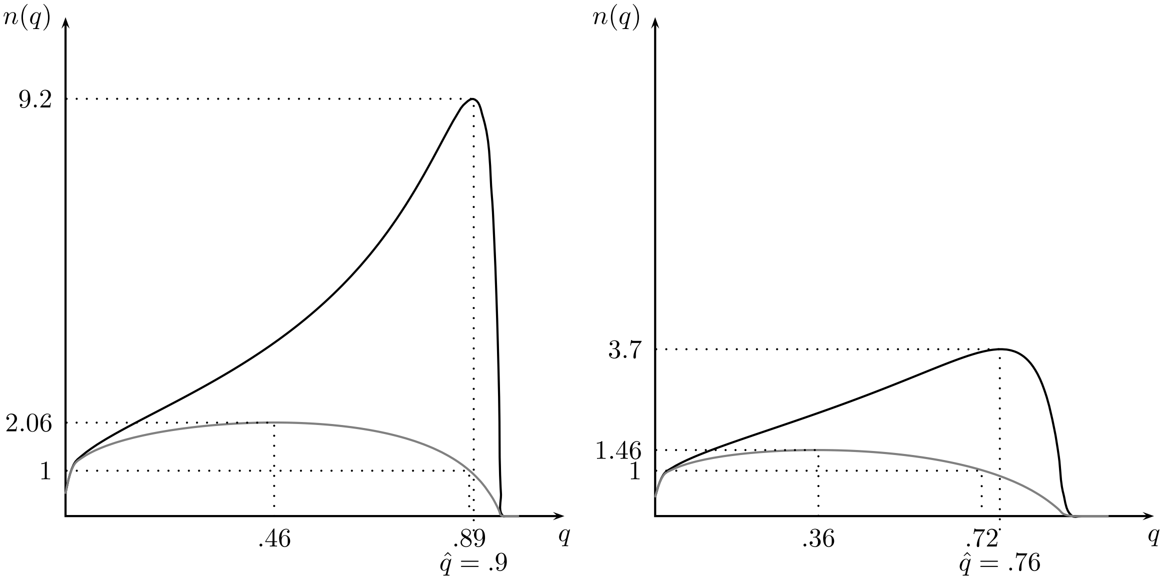
Differentiating (10) with respect to p^S gives

$$\frac{\partial^2 n^*}{\partial q \partial p^S} = \frac{\partial^2 n^*}{\partial q \partial A}\frac{\partial A}{\partial p^S} = -\frac{1}{(c^{spam}+p^S)\ln(q)^2 q} < 0. \tag{14}$$

This shows that the larger is p^S, the slower is the initial increase in the number of variant messages sent to each target as q increases from zero and the smaller is the maximum n^*. Again, a similar result holds for p^R:

$$\frac{\partial^2 n^*}{\partial q \partial p^R} = \frac{\partial^2 n^*}{\partial q \partial A}\frac{\partial A}{\partial p^R} = -\frac{f(p^R)}{ln(q)^2 q(1-F(p^R))} < 0. \tag{15}$$

The flavor of these results are illustrated in Figures 1a and 1b. The functions $n^*(q)$ and $n^{inbox}(q)$ are illustrated for $A = 0.04$ in Figure 1a and for $A = 0.1$ in Figure 1b. When $A = 0.04$ the expected revenue of spam is relatively large compared to the spammer's marginal cost of sending an additional message. Increasing the effectiveness of the filter from $q = 0$ in this case causes the spammer to increase the number of messages sent from approximately one per target to $n^* = 9.2$ per target at $\hat{q} = 0.9$. Any further improvements in the effectiveness of the filter will lead to a rapid decline in the number of messages sent per target getting n^* below 1 (approximately the pre-filtering volume) at $q = 0.96$. Furthermore, n^{inbox} reaches its maximum 2.06 messages per target when $q = 0.46$ and remains above one (approximately the pre-filtering volume) for $q \leq 0.89$. In other words, unless filters are very effective, their introduction will lead to both a rapid increase in messages sent per target and an increase in the number of messages arriving in a target's inbox. When the expected revenues of spam are smaller compared to the spammer's marginal cost of sending an additional message the potential problems associated with filters are not as severe. With $A = 0.1$, n^* peaks at 3.7 where $\hat{q} = .76$, $n^* = 1$ when $q = .89$ and n^{inbox} peaks at 1.46 and remains above one for $q \leq 0.72$. As A is increasing in p^S and p^R, these results imply that when sender or receiver pricing is used in conjunction with filtering, filtering is less likely to lead to a massive increase in messages per target as filters become more effective. Furthermore, the higher are the sender and/or receiver prices the more likely we are in the range where filters reduce the volume of spam below the no-filter benchmark.



a: $A = 0.04$ b: $A = 0.1$

Figure 1: $n^*(q)$ in black and $n^{inbox}(q)$ in gray

Surely these results must cast doubt on the ability of filtering alone to solve the spam problem suggesting instead that filters have the potential to make the problem worse both in terms of the number of messages being sent in total and the expected number of messages arriving in an individual target's inbox.

It is clear that because filtering improves the ability of prices to reduce spam, the spam-eliminating prices are the lower the more effective is the filter. We can show this for sender price by differentiating (7) with respect to q.

$$\frac{\partial p^S_{spam}}{\partial q} = -\frac{(1 - F(p^R))\pi}{q} < 0. \tag{16}$$

and for the receiver price by differentiating (6) with respect to q:

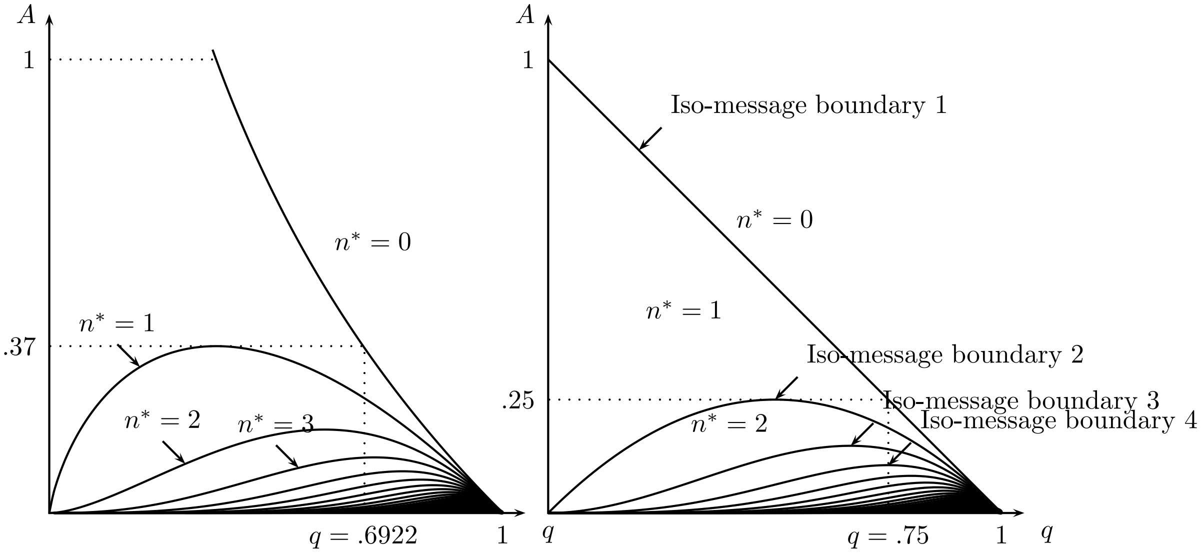
$$\frac{\partial p^R_{spam}}{\partial q} = -\frac{(c^{spam} + p^S)}{q(ln(q))^2\pi^2 f(p^R_{spam})} < 0 \tag{17}$$

The spam-eliminating receiver price decreases with p^S and the spam-eliminating sender price decreases with p^R, which suggests that the receiver pays price and the sender pays price could be used in conjunction with each other to deter spam.

$$\frac{\partial p^S_{spam}}{\partial p^R_{spam}} = f(p^R)\pi ln(q) < 0. \tag{18}$$

The above analysis of the interplay between the prices that affect A, the level of filter q and the spammer's choice of the number of messages per target n^* is illustrated in Figure 2. The left-hand side figure is the illustration of the spammer's problem in the continuous choice model given above. The iso-message curves are obtained by setting n^* in (2) equal to discrete values $\{0, 1, 2, ...\}$. The graph can be read in multiple ways. First, we can see the spammer's choice of n^* as the effectiveness of the filter varies for given level of prices and thus A. Increasing q from approximately zero towards one will first increase n^* before n^* starts to decline. This is true for all levels of A. The essence of this relationship was also shown in Figure 1. Second, for a given filter effectiveness, we can see that increases in the prices, and thus an increase in A, lead to a reduction in n^* and that this reduction is the faster the larger is q.

As discussed briefly above, the continuous form of the model is not defined at $q = 0$, and the intuitive result that $n^* = 0$ when $q = 0$ is not generated in this model. However, if we look at the discrete form of the model, given on the right-hand side, some intuitively appealing properties of the model return. This graph shows the iso-message regions where exactly 1,2,3,... messages are sent per target as well as the boundaries of these regions. Thus, an iso-message boundary is the locus of points where the spammer is indifferent between sending $n-1$ and n messages per target. We can see that $n^* = 1$ if $q = 0$ and if $A \leq 1$. Increasing the effectiveness of the filter does not affect the spammer's choice if strictly inside the iso-message boundaries. For example, if $q > .25$, $n^* = 1$ as long as $q < 1 - A$, and any further improvements in q will lead to the elimination of spam. Notice that in this area, $n^{inbox} = (1-q)$ reduces linearly with q. However, the properties of the continuous model, particularly the result that increases in q first increase n^* before reducing it, are seen when $A < .25$. When $A < 0.25$, each targeted consumer can expect to receive more than one message in their inbox if $q < \frac{(n-1)}{n}$ and less than one message in their inbox if $q > \frac{(n-1)}{n}$. Clearly, if $q < \frac{(n-1)}{n}$ consumers are worse off in terms of the number of spam messages they receive in the presence of filtering than they would be in the absence of filtering because of the perverse incentives that filtering provides to the spammer. For example, from equations (34) and (35) we see that if $A = 0.1$ a filter that blocks 50% of all spam will result in 3 messages being sent by the spammer to each consumer on his list and each of these consumers can expect to receive 1.5 messages in their inbox. The discrete representation of the model is derived fully in the Appendix.



a: Continuous model

b: Discrete model

Figure 2: Iso-message curves for the continuous model and iso-message boundaries and regions for the discrete model

To conclude, we have shown that prices and filtering work best when used together and are therefore complements in the war against spam.

4 Stage 1 - Profit maximizing size of mailing list

In stage 1, the spammer chooses the size of his mailing list N at a total cost $C(N)$ where the usual cost function properties $C'(N) > 0$ and $C''(N) > 0$ are assumed to hold. The stage 1 objective function for the spammer is

$$V \equiv \Pi(n^*)N - C(N), \tag{19}$$

where $\Pi(n^*)$ is the expected stage 2 profit per target in (5). The optimal size of the mailing list N^* is implicitly defined by the equilibrium condition

$$\Pi(n^*) \equiv C'(N^*) \tag{20}$$

if an interior solutions exists, that is, if $\Pi(n^*) > C'(0)$.

The assumed cost function properties are quite intuitive. Spammers acquire addresses using several methods. They may simply purchase address lists. It is conceivable that the more lists they have, the less likely it is that a new list will generate new unique names, and thus given a constant cost per list and given that the spammer purchases a large number of lists, this method is consistent with an increasing marginal cost of a unique name. Second, they may cull addresses from the internet using web-crawling software. The time-cost of a new unique name goes up with the number of unique names already found, and again we can expect there to be an increasing marginal cost of a unique name.

Eaton, MacDonald and Meriluoto (2007) generate a specific cost function by assuming that the advertiser uses random sampling with replacement from population H with a constant cost of a draw v to build a mailing list. Given that the number of unique names on the list is N, this technology yields a marginal cost of generating a unique address as

$$C'(N^*) = \frac{vH}{H-N} \tag{21}$$

that has the properties assumed above. We will not need to use this specific cost function in the comparative static results below because more general results are easy to obtain.

Let the total number of messages sent by the spammer be

$$T = n^* N^*. \tag{22}$$

4.1 Comparative statics on N^*

Totally differentiating (20) yields

$$C''(N^*)dN = d\Pi(n^*) \tag{23}$$

and thus we can see that

$$\frac{\partial N^*}{\partial \Pi(n^*)} = \frac{1}{C''(N^*)} > 0. \tag{24}$$

Thus, any activity that leads to a reduction in the spammers profit per target will also reduce the size of the spammer's mailing list. Because we know that not only sender and receiver prices but also filtering reduce the spammer's equilibrium profit, we can conclude that the optimal size of the spammers' mailing list is inversely related to the sender price, the receiver price and the effectiveness of the filter.

4.2 Comparative statics on T

The total volume of spam varies with p^R, p^S and q in the following way:

$$\frac{\partial T}{\partial p^R} = \frac{\partial n^*}{\partial p^R} N^* + \frac{\partial N^*}{\partial p^R} n^* < 0 \tag{25}$$

$$\frac{\partial T}{\partial p^S} = \frac{\partial n^*}{\partial p^S} N^* + \frac{\partial N^*}{\partial p^S} n^* < 0 \tag{26}$$

and

$$\frac{\partial T}{\partial q} = \frac{\partial n^*}{\partial q} N^* + \frac{\partial N^*}{\partial q} n^*. \tag{27}$$

Given (25) and (26) it is clear that increasing the receiver price and/or the sender price unambiguously reduce the total volume of spam. The impact of filtering on the total volume of spam in (27), however, depends on the value of q. From equations (10), we know that for $q > \hat{q}$ the number of messages per target is a decreasing function of q and from (24) we can see that the optimal mailing list decreases with q and so (27) is negative over this range. For $q < \hat{q}$, however, (10) is positive and so without pinning down parameter values, we cannot comment on whether the effect on the size of the mailing list outweighs the effect on the number of messages sent to each target. We do know for certain though that as q approaches $\hat{q}$ from below, N^* starts to decline before n^* does.

5 Welfare implications of controlling spam

Our understanding of how filtering and prices affect the number of messages sent to each target allows us to partially address the issue of how efforts to control spam might impact on social welfare. It is tempting in the face of anecdotal evidence to simply assume that spam is a sufficiently large problem that it should be eliminated at all costs. In reality, of course, things are not so clear-cut and one must be concerned with balancing the benefits of less spam, if they exist, with the costs that filters and prices impose on other aspects of the email network. Unless spam controls can be targeted specifically at spam and spam alone, which seems very unlikely to be possible, we will need to take into account the unintended impacts of spam control on non-spam email.

In what follows we present a spam-specific welfare function and discuss, one by one, the ways that filters and prices affect it. We look both at controls that are discriminatory in that they target only spam, thereby allowing us to ignore the impacts of pricing on non-spam welfare, and controls that are non-discriminatory. With the non-discriminatory controls, we need to incorporate the welfare of non-spam messages. To do that, we use an analysis similar to those of Hermalin and Katz (2004) and Loder et. al. (2006) with the exception that we consider spam messages separately from the non-spam messages.

A complete analysis of the welfare effects of controlling spam would also need to take into account the cost of introducing controls but we ignore these here. Experience has already shown us that maintaining a filtering system is not cheap and the cost of designing and implementing a pricing mechanism might be similarly expensive. However, without empirical data to pin down parameter values in our theoretical model we cannot say anything concrete in this regard.

5.1 Spam-sector welfare

The expected social welfare of sending n spam messages to a single target is

$$w^{spam} = (1-F(p^R))(1-q^{n(q,p^S,p^R)}) \left(\pi + \frac{\int_{p^R}^{\rho^{\max}} \rho f(\rho) d\rho}{1-F(p^R)} \right) - (c^S + c^U + (1-q)c^R) n(q,p^S,p^R) \tag{28}$$

and the total social welfare of spam is

$$W^{spam} = w^{spam} N(q,p^S,p^R) - C(N(q,p^S,p^R)). \tag{29}$$

The expected social benefit of spam is made up of both the spammer's expected profit and the benefit of the recipient if she chooses to read it. The social cost of a single spam message is simply the sum of spammer's, ISP's and recipient's per message costs.

Clearly the objectives of the spammer and society are not perfectly aligned and so, in the absence of controls, the spammer's choices N^* and n^* are unlikely to be equal to the welfare maximizing N^{opt} and n^{opt}. That said, the divergence in objectives arises with respect to n in that if controls can be put in place that internalize all per message external effects then $n^* = n^{opt}$, $\Pi(n^*) = w^{spam}$ and so $V = W^{spam}$. In other words, if the spammer can be made to choose $n^* = n^{opt}$ then they will also choose $N^* = N^{opt}$.

5.2 Discriminatory controls

Here we analyze the welfare effects of a sender price, a receiver price and filtering that target spam only. The primary effect of these controls is to influence the spammer's choice of n^* and we have shown that higher prices unambiguously lead to fewer spam messages and filters can lead to either an increase or decrease in the number of spam messages per target depending on their effectiveness. A secondary, and with respect to social welfare unintended, effect of receiver prices and filters is to reduce the expected benefit that spam lovers receive from spam either because the messages that are sent are not received or read. The use of a sender price has no such secondary effect.

If the goal of society is to completely eliminate spam ($n^{opt} = 0$), any of the three controls can be equally effective. The spam-eliminating receiver price is implicitly found in (6), the spam-eliminating sender price is found in (7) and the required filter effectiveness is implicitly found in (2).

If the goal of society is not to eliminate spam but to reduce its amount ($0 < n^{opt} < n^*$), the best control to use is a sender price because it does not have the welfare reducing secondary effect on those who like spam.

If the goal of society is to increase the amount of spam ($n^{opt} > n^*$) and currently no controls are in use, then only a sender price $-c^{spam} \leq p^S < 0$ can be used.

5.3 Non-discriminatory controls

In reality prices cannot be levied only on spammers and filtering false captures some amount of non-spam email. While this means that it is possible that the welfare gains resulting from the reduction of spam will be partially or fully offset by a loss in non-spam email welfare, it should be noted that zero prices for non-spam email are suboptimal and spam-eliminating prices are likely to be small.

We assume that net benefits associated with a non-spam email message are captured by the pair (σ, ρ), where σ is the benefit the sender gets if the message is read, and ρ is the benefit the receiver gets from reading the message. Both σ and ρ can be positive, negative, or zero. Messages that are sent and read give the sender and receiver the following private surpluses:

$$s^S = \sigma - c^S - p^S \tag{30}$$

and

$$s^R = \rho - c^R - p^R. \tag{31}$$

The social welfare of a non-spam message is

$$w^{non-spam} = \sigma + \rho - C, \tag{32}$$

where $C = c^S + c^U + c^R$.

Given (p^S, p^R), messages in the following set will be exchanged

$$SR(p^S, p^R) \equiv \{(\sigma, \rho) | \sigma \geq \max(c^S + p^S, 0), \rho \geq \max(p^R, 0)\}. \tag{33}$$

Figure 3 illustrates a possible distribution of preferences (without any consideration for density) in the (σ, ρ)-space. All messages above the line $\rho = C - \sigma$ are welfare-improving and all messages below the line are welfare-reducing. With the introduction of any arbitrary uniform prices p^S and p^R there are two effects on welfare. First, inefficient messages in $abefgja$ are not exchanged and this improves welfare. Second, efficient messages in $bcde$ and $ghij$ are not exchanged and this reduces welfare. Socially optimal prices must therefore balance these opposing welfare effects at the margin.

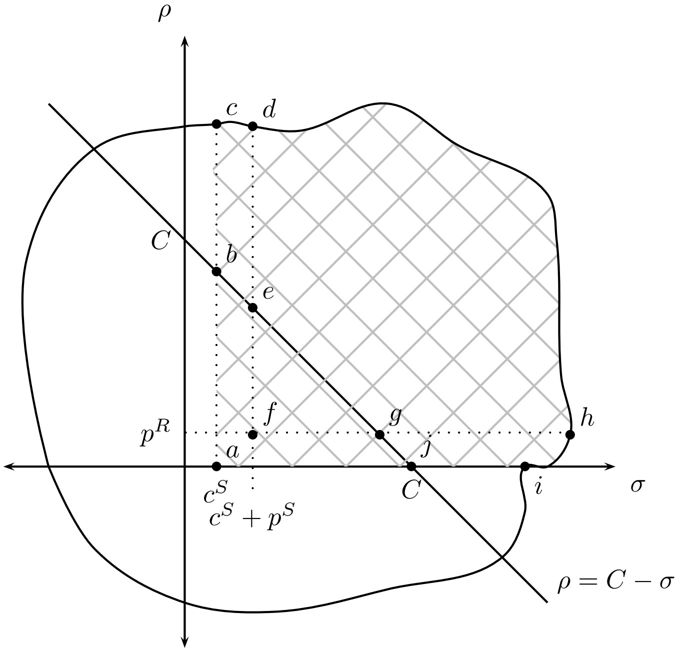


Figure 3: Messages that are sent and read given zero prices

If either of the spam-eliminating prices is smaller than the optimal uniform prices for non-spam email, and if the goal of society is to completely eliminate spam ($n^{opt} = 0$), then the use of the optimal uniform prices for non-spam email will eliminate spam and maximize total social welfare. If, however, both the spam-eliminating prices are larger than the optimal uniform prices for non-spam email then a more detailed cost-benefit analysis needs to be undertaken to identify what mix of prices and filtering maximizes total social welfare and whether this mix completely eliminates spam or not. Clearly the optimal mix depends in part on the distribution of non-spam messages. Other things equal, the use of a sender price to control spam will have less associated detrimental non-spam impact if there are relatively few messages that generate little value for senders but lots of value for receivers (region *bcde* in Figure 3). The use of a receiver price will have less associated detrimental non-spam impact when there are relatively few messages that would generate large value for senders but little value for receivers (region $ghij$ in Figure 3).

Finally, recall that increasing the effectiveness of filters reduces the magnitude of both spam-eliminating prices and so it is possible, in principle, to have a mix of controls that ensures that spam can be eliminated by using prices that maximize welfare in the non-spam network. However, the benefit of increasing q must be weighed against the likelihood of increasing the cost of falsely filtering non-spam messages.

6 Conclusion

We have examined receiver pays pricing, sender pays pricing and filtering solutions to the spam problem. Receiver pays pricing works by reducing the incentive of the receivers of spam messages to open them and sender pays pricing works by increasing the spammer's costs. Filtering alone is unlikely to offer a viable solution to the spam problem if spammers counteract it by sending multiple variants of a message to each consumer. In fact, our model suggests that filtering can be counterproductive by leading to an increase in the total volume of spam and sometimes even in the number of spam messages arriving in consumers' inboxes. However, we show that the more effective is the filter, the more effective are receiver and sender prices in reducing spam and the lower in magnitude are the spam-eliminating prices. Both these results imply that the potential welfare loss of pricing, due to a loss in 'good' messages in the consumer to consumer network, would be minimized with effective filtering. The prices and filtering are therefore complements in the war against spam.

A A discrete representation of spammer behavior

If the spammer is restricted to choose the number of messages to send, $n \in \{0, 1, ..., \infty\}$, the spammer sends no messages ($n^* = 0$) if $\Pi(1) < 0$, sends one message ($n^* = 1$) if $\Pi(1) \geq 0$ and if $\Pi(1) > \Pi(2)$, etc. Generally, $n^* = n$ if

$$q^n(1-q) < A \leq q^{n-1}(1-q). \tag{34}$$

The expected number of messages that actually make it into the inbox of a consumer on the spammer's mailing list is:

$$n^{inbox} = (1-q)n^*(q). \tag{35}$$

If $q = 0$ the spammer sends at most one message to each consumer on his list and all of these messages arrive in the consumers' inboxes. When $q > 0$ and $A > 0.25$, the spammer sends at most one spam message to each consumer on his mailing list and the expected number of messages received by each targeted consumer is $(1-q)$ if $q < 1 - A$ and zero if $q \geq 1 - A$. For $q > 0$ and $A < 0.25$, the discrete model and the continuous model behave similarly and so the flavor of the comparative static results derived in equations (8)-(18) is evident in Figure 2b for this region of the parameter space.

If $A < 0.25$, there is a range of values for q such that the spammer sends two or more messages to each consumer on his mailing list. For very small A the volume of spam increases rapidly as q increases from zero and starts to decrease only when q takes on values close to one. When $A < 0.25$, each targeted consumer can expect to receive more than one message in their inbox if $q < \frac{(n-1)}{n}$ and less than one message in their inbox if $q > \frac{(n-1)}{n}$. Clearly, if $q < \frac{(n-1)}{n}$ consumers are worse off in terms of the number of spam messages they receive in the presence of filtering than they would be in the absence of filtering because of the perverse incentives that filtering provides to the spammer.

The spam eliminating price in the discrete representation of the model, $p^{S'}_{spam}$, is

$$p^{S'}_{spam} = \alpha\pi(1-q) - c^{spam} \tag{36}$$

and it decreases linearly in the effectiveness of the filter.

References

[1] Arrison, Sonia (2004) "Canning Spam: An economic solution to unwanted email". *Pacific Research Institute study.*

[2] Dai, R and Li, K. (2004) "Shall we stop all unsolicited email messages?" *Proceedings of First Conference on Email and Anti-Spam (CEAS).*

[3] Fahlman, S. (2002) "Selling interrupt rights: a way to control unwanted e-mail and telephone calls." *IBM Systems Journal* 41(4), 759-66.

[4] Griffiths, P. (2006) "Email gangs bombard web in 'spam wars"' *The Press*, Christchurch, 29 November 2006.

[5] Hermalin, B. E. and M. L. Katz (2004) "Sender or receiver: who should pay to exhcange an electronic message?" *RAND Journal of Economics* 35(3), 423-448.

[6] Howell, D. (2004) "E-mail spammers need only a few customers to get meaty rewards." *Investor's Business Daily*, 19 February 2004.

[7] Khong, D. (2004) "An economic analysis of spam laws", *Erasmus Law and Economics Review* 1, 23-45.

[8] Kraut, R., Sunder, S., Telang, R. and J. Morris (2005) "Pricing electronic mail to solve the problem of spam", *Human-computer Interaction* 20, 195-223.

[9] Loder, T., Van Alstyne, M. and R. Wash (2006) "An Economic Response to Unsolicited Communication," *Advances in Economic Analysis & Policy* 6(1), Article 2.

[10] The Economist (2007) "Spam seems here to stay." *Seattle Post - Intellegencer*, September 28, 2007.

[11] US Public Law (2003), "Congressional Findings and Policy." *Can-Spam Act of 2003*, US Public Law No. 108-187, 117 Stat., Sec. 2, December 16, 2003.

[12] Van Alstyne, M. (2007) "Curing spam: rights, signals & screens." *Economists' Voice*, 4(2), March 2007.