

Spam - solutions and their problems

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

January 31, 2008

Abstract

We analyze three potential solutions to the spam problem - sender pays pricing, receiver pays pricing and filtering - used alone or concurrently. We find that filters alone may exacerbate the spam problem if the spammer tries to evade them by sending multiple variants of the message to each consumer. Sender pays and receiver pays prices can be effective on their own or with filtering in reducing or eliminating spam. When filtering is used in conjunction with either price the magnitude of the spam-eliminating price is unambiguously reduced for every level of filter effectiveness.

Keywords: spam, filtering, email, receiver pays pricing, sender pays 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 and the 2007 E.A.R.I.E. Conference 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 systems¹, 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. Even though few people who are contacted by spammers are interested in the products on offer, and spammers are incapable of identifying who these potential customers are ex ante, the fact that the marginal cost of sending an email message is extremely small implies that the expected benefit of a spam message need not be large for spam to be profitable. 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,

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

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).

A number of technological defenses designed to filter or block unwanted messages from consumers' inboxes are also available to both ISPs and consumers but these too have proven to be ineffective at eliminating the spam problem. Blacklisting blocks messages sent by specific senders who have been identified as undesirable. Whitelisting blocks all messages except those coming from specific senders identified as acceptable. 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 false negatives and avoiding false positives. With blacklisting there is essentially no chance of a false positive but the process is completely ineffective if spammers can easily hide or quickly change their identities. Whitelisting is more effective at eliminating false negatives but eliminates the scope for email to be used as a means of communication with people outside of one's immediate sphere of contacts. Filters can be adjusted to control the balance between 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 the blocked messages and so only some of the costs of spam are eliminated. Moreover, if spammers can increase the likelihood of evading filters by sending multiple variants of a message to each consumer it is entirely possible that these technical 'solutions' actually exacerbate the problem of spam by increasing both the total volume of spam and the number of spam messages that arrive in consumers' inboxes.

Economic defences against spam discussed in the literature include sender pays 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 pays 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 as well as sender and receiver pays pricing to 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 pays pricing could reduce or eradicate spam by reducing the number of consumers who will read spam messages therefore reducing the expected marginal benefit of sending spam. Similarly sender pays pricing could reduce or eradicate spam by reducing the spammer's expected profit per customer. 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, but that the potential problems associated with filtering are reduced when used together with either sender pays and/or receiver pays pricing.

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 pays and sender pays prices and the effectiveness of the filter. We find that the magnitude of the spam-eliminating receiver pays and sender pays prices are inversely related to the effectiveness of the filter suggesting that filters and prices complement each other in the fight against spam.

Two practical issues associated with implementing prices are discussed in Section 5. First, in the formal model there is a single ISP offering email services to all consumers and the monopoly spammer. This ISP uses a filter and may charge sender pays and/or receiver pays prices. In the real world, however, consumers and spammers can choose from any number of ISPs who might be following different pricing policies and so a consideration of ISP competition is important. If, in an effort to get a competitive edge, ISPs have an incentive to lower email prices their effective adoption might require cooperation or regulation. We also discuss the practical problem of implementing pricing that arises when spammers and/or consumers are able to avoid having to pay the prices, rendering pricing ineffective.

Loder, Van Alstyne and Wash (2006) analyze the welfare effects of three competing economic responses to unsolicited email: flat tax, perfect filter and attention bond. The filter discussed by Loder et. al. is a perfect filter that eliminates all unwanted messages (those with value less than the processing cost to

the recipient). Thus, all wanted spam messages that are sent would still get through to the consumers whereas all unwanted spam would be blocked by the filter. From the spammer's viewpoint, fewer messages will be read due to the filter and so the expected value of a message falls² which in turn reduces the number of messages sent by the spammer. The filter thus leads to an unambiguous reduction in total volume of spam. Furthermore, the number of unwanted spam messages in consumers' inboxes goes to zero while the consumers who like spam receive a reduced number of spam messages. Our filter, in contrast, blocks any spam message with probability q . All consumers, therefore, receive any particular spam message with probability $(1 - q)$ regardless of their tastes for spam. The spammer's likelihood of getting at least one message through the filter when sending a total of n messages is $(1 - q^n)$ and so unless the filter is very good, it exacerbates the spam problem by inducing the spammer to send multiple variants of a message. This implies that the volume of spam is likely to go up with filtering, a result in contrast to that of Loder et. al. A casual observation about the explosion of spam in the last few years when filters have become the norm seems to support our view. Finally, because the way we treat filtering in our model differs from that of Loder et. al., filters and sender pays prices turn out to be complements in combatting spam and not substitutes as they are in Loder et. al.

The flat tax analyzed by Loder et. al. is equivalent to our sender pays price. They assume that a single governmental agency can impose taxes on all senders of messages in a network. Thus, the tax can be designed to internalize the external effect of a message to the receiver, also on the same network. The setup is similar to our monopolist spam model where there is a single ISP serving the network. Our analysis is not focused on issues of efficiency but rather how

²This 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.

the prices affect the volume of spam, the levels of the spam-eliminating prices and how the prices and filtering can work together to combat spam.

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 target and the number of messages that each target receives in her inbox in Section ???. 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. Section 5 discusses various technical and/or coordination problems associated with our solutions to spam. This section is qualitative with no specific theoretical model in mind. 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 modeling ISP decisions per se, we treat the ISP as if it were a single autonomous entity that services all participants in the email network. We leave it to Section 5 to discuss how competitive ISP behavior might influence spam outcomes.

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 message³. 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 send⁴. This per message cost for the spammer is certainly small and likely to be very close to zero. Each spam message costs the ISP c^U to transmit and each message that arrives in a consumer's inbox costs that consumer c^R to process. The spammer pays a per message sender pays price $p^S \geq 0$ to the ISP⁵ and the receiver of a message pays a per message receiver pays price of $p^R \geq 0$ to the ISP upon opening a message. We place two restrictions on p^R in order to rule out receiver pays 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

³We 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.

⁴In 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.

⁵In the absence of price discrimination consumers also pay this price to send messages but we ignore this detail as we do not explicitly model consumer welfare in this paper.

pays price for those messages in their inbox that they choose to open.

If consumer i reads a spam message she receives utility ρ_i^{spam} , drawn from a smooth and continuous distribution in $[\rho_{min}^{spam} < 0, \rho_{max}^{spam} > 0]$ with positive density for all ρ_i^{spam} in the range. We assume that the heading and the sender information contain enough information about a message for the consumer to infer its value. We assume that the proportion α of all consumers who receive positive ρ_i^{spam} purchase the spammer's product if they read his message. Positive values of ρ_i^{spam} are associated with gaining valuable product information and reduced search costs.

Given the restrictions on p^R and the assumptions on the distribution of ρ_i^{spam} above, it is clear that consumers with $\rho_i^{spam} \leq 0$ will never read the spam message and that the proportion of spam lovers who read the spam message is a smooth, continuous and decreasing function of p^R : $\theta = \theta(p^R)$ with $\frac{\partial \theta(p^R)}{\partial p^R} < 0$. Because the processing cost of the receiver, c^R , is sunk at the time of the decision to read a message, $\theta(0) = 1$, that is all interested consumers read the spam message when $p^R = 0$. We also know that $\theta(\rho_{max}^{spam}) = 0$ and that $0 < \theta(p^R) < 1$ for any price in $0 < p^R < \rho_{max}^{spam}$.

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 pays 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.

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, by drawing with replacement from a population of N addresses. 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 . We use backward induction to solve the problem in the next two subsections and to determine how the spammer's choices are affected by receiver pays pricing, sender pays pricing and filtering.

3 Stage 2 - Optimal 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 repre-

sentation of the spammer's per target message choice in subsection 3.3. Due to being well-defined at $q = 0$, this discrete version of the problem has a more intuitive graphical interpretation than the continuous version.

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

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

The profit-maximising number of messages per target is

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

where $A = \frac{c^{spam} + p^S}{\theta(p^R)\alpha\pi}$ is the ratio of the spammer's marginal cost and expected marginal revenue. The expected number of messages received by each targeted consumer is

$$n^{inbox} = \begin{cases} \frac{(1-q)\ln(-\frac{A}{\ln(q)})}{\ln(q)} & \text{if } A \leq -\ln(q) \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The maximized profit equals

$$\Pi(n^*) = \begin{cases} (1 - q^{\frac{\ln(-\frac{A}{\ln(q)})}{\ln(q)})}\theta(p^R)\alpha\pi - \frac{(c^{spam} + p^S)\ln(-\frac{A}{\ln(q)})}{\ln(q)} & \text{if } A \leq -\ln(q) \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

It is easy to show, for future reference, that $\frac{\partial \Pi(n^*)}{\partial q} < 0$.

3.1 Spam-eliminating p^R and p^S

Define $\underline{\theta}$ as the level of θ that solves $n^* = 0$:

$$\underline{\theta} \equiv -\frac{c^{spam} + p^S}{\alpha\pi\ln(q)}. \quad (5)$$

$\underline{\theta}$ is the lowest proportion of interested consumers who open the spammer's message required for spamming to be profitable. $\underline{\theta}$ is increasing in q , c^{spam} and

p^S that lower the spammer's expected profit and decreasing in α and π that increase the spammer's expected profit. The spam-eliminating receiver pays price p_{spam}^R is implicitly defined as

$$\theta(p_{spam}^R) \equiv \underline{\theta}. \quad (6)$$

Notice that $p_{spam}^R \leq \rho_{max}^{spam}$.

The spam-eliminating sender pays price is

$$p_{spam}^S = -\theta(p^R)\alpha\pi\ln(q) - c^S. \quad (7)$$

3.2 Comparative statics

In this section, we will 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 way:

$$\frac{\partial n^*}{\partial p^R} = \frac{\partial n^*}{\partial A} \frac{\partial A}{\partial \theta} \frac{\partial \theta(p^R)}{\partial p^R} = -\frac{1}{\theta(p^R)\ln(q)} \frac{\partial \theta(p^R)}{\partial p^R} \leq 0 \quad (8)$$

because $\frac{\partial \theta(p^R)}{\partial p^R} \leq 0$,

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

and

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

It is clear from (8) and (9) that both receiver and sender pays pricing unambiguously deters spam but the impact of the effectiveness of the filter on the number of messages per target in (10) is ambiguous. Defining $\hat{q} \equiv e^{-Ae}$ as the value of q that sets $\frac{\partial n^*}{\partial q} = 0$, we see from (10) that if $q < \hat{q}$, n^* increases as q increases and if $q \geq \hat{q}$, n^* decreases as q increases.

From (10), it is evident that receiver and sender pays prices also have an indirect effect on spam through filtering. Differentiating \hat{q} with respect to p^S yields

$$\frac{\partial \hat{q}}{\partial p^S} = \frac{\partial \hat{q}}{\partial A} \frac{\partial A}{\partial p^S} = -\frac{1}{\theta(p^R)\alpha\pi} e^{-Ae+1} < 0. \quad (11)$$

and differentiating (10) with respect to p^S gives

$$\frac{\partial^2 n^*}{\partial q \partial p^S} = \frac{\partial^2 n^*}{\partial q \partial A} \frac{\partial A}{\partial p^S} = -\frac{1}{(c^{spam} + p^S) \ln(q)^2 q} < 0. \quad (12)$$

These results show 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, the lower is the cut-off value \hat{q} beyond which increases in q result in a reduction in the number of messages sent to each consumer and the smaller is the maximum n^* . As $sign\left(\frac{\partial A}{\partial p^S}\right) = sign\left(\frac{\partial A}{\partial \theta} \frac{\partial \theta(p^R)}{\partial p^R}\right)$, equivalent comparative static results as given in (11) and (12) hold for p^R . These general results are illustrated in Figures 1a and 1b below. These results imply that when sender pays pricing or receiver pays pricing is used in conjunction with filtering, the potential problems of filtering are reduced.

The functions $n^*(q)$ and $n^{inbox}(q)$ are illustrated for $A = 0.04$ in Figure 1a and for $A = 0.1$ in in Figure 1b. When $A = 0.04$, the number of messages sent to each target n^* peaks at 9.2 when $\hat{q} = .9$. Furthermore, n^{inbox} reaches its maximum 2.06 messages per target when $\hat{q} = .46$ and remains above one for $q \leq 0.89$. When $A = 0.1$, n^* peaks at 3.7 when $\hat{q} = .76$ and n^{inbox} peaks at 1.46 and remains above one for $q \leq 0.72$. 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 rather than better, 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.

The spam-eliminating sender pays price and the spam-eliminating receiver

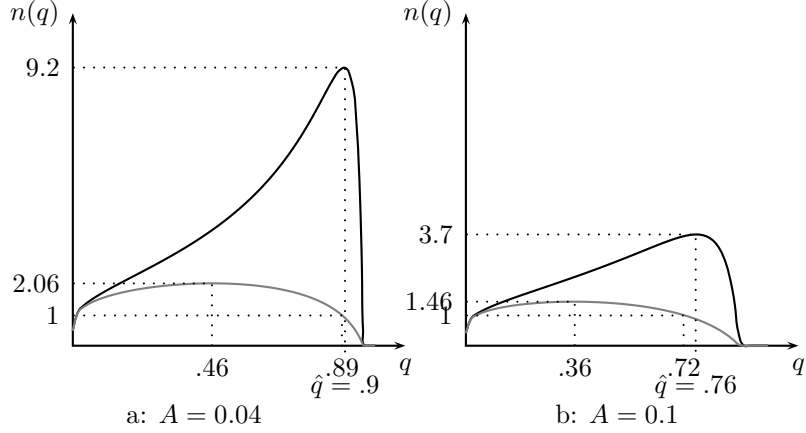


Figure 1: $n^*(q)$ in black and $n^{inbox}(q)$ in gray

pays price decrease as q increases as is shown in equations (13) and (14):

$$\frac{\partial p_{spam}^S}{\partial q} = -\frac{\theta(p^R)\alpha\pi}{q} < 0 \quad (13)$$

and

$$\text{sign}\left(\frac{\partial p_{spam}^R}{\partial q}\right) = -\text{sign}\left(\frac{\partial\theta}{\partial q} = \frac{c^{spam} + p^S}{\alpha\pi\ln(q)^2q}\right) < 0. \quad (14)$$

These results are potentially very important. While we have not engaged in welfare analysis in this paper, it is worth noting that any pricing solution would not only lead to a reduction in spam but also in other email traffic. If these lost emails were welfare-improving, then some of the benefit from the reduction in spam levels would be lost due to the loss of these “good” messages. Therefore, the smaller is the spam-eliminating price the fewer other messages will be lost and therefore the lower is the welfare cost.

The spam-eliminating receiver pays price decreases with p^S and the spam-eliminating sender pays 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.

$$\frac{\partial p_{spam}^S}{\partial p_{spam}^R} = -\alpha\pi\ln(q)\frac{\partial\theta(p^R)}{\partial p^R} < 0. \quad (15)$$

Since we have not engaged in welfare analysis, we are not in a position to say if there is an optimal combination of sender and receiver pays prices.

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

3.3 A discrete representation of spammer behavior

If the spammer is restricted to choose the number of messages to send, $n \in \{0, 1, \dots, \infty\}$, 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

$$q^n(1 - q) < A \leq q^{n-1}(1 - q). \quad (16)$$

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

$$n^{inbox} = (1 - q)n^*(q). \quad (17)$$

We illustrate the combinations of A and q for which the spammer chooses to send exactly n messages in an iso-message region mapping in Figure 2. The two inequalities in (16) define the iso-message region n bounded from below by iso-message boundary $n + 1$ and from above by iso-message boundary n .

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)-(15) is evident in Figure 2 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

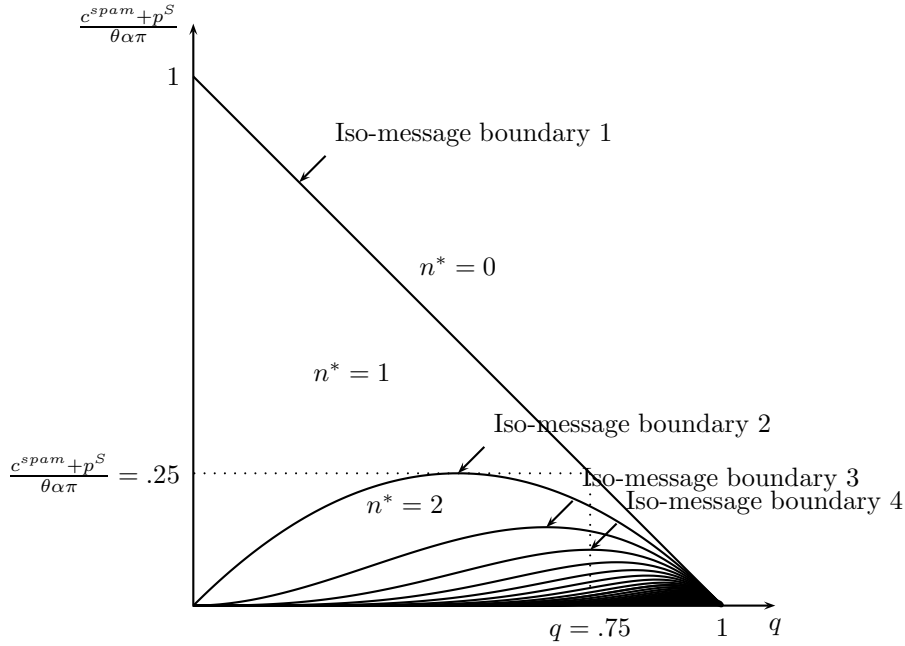


Figure 2: Iso-message regions and boundaries in a discrete representation of the model

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. For example, from equations (16) and (17) 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 spam eliminating price in the discrete representation of the model,

$p_{spam}^{S'}$, is

$$p_{spam}^{S'} = \alpha\pi(1 - q) - c^{spam} \quad (18)$$

and it decreases linearly in the effectiveness of the filter.

4 Stage 1 - Optimal size of mailing list

In stage 1, the spammer chooses the size of its mailing list N at a total cost $C(N)$. The stage 1 objective function for the spammer is

$$V \equiv N\Pi(n^*) - C(N), \quad (19)$$

where $\Pi(n^*)$ is the expected stage 2 profit per target in (4). The equilibrium condition is simply

$$\Pi(n^*) = C'(N) \quad (20)$$

if an interior solutions exists, that is, if $\Pi(0) > 0$.

We generate a specific cost function by assuming that the spammer builds his mailing list by drawing addresses with replacement from the entire population H of email users. Let the cost of a draw be v . If the spammer already has a sample of size N , the probability of getting a unique address in the next draw is $\frac{H-N}{H}$ and, since the spammer is drawing with replacement, this probability remains constant until he has found an additional unique address. The expected number of draws, x , required to find one more unique address is therefore found from $x\frac{H-N}{H} = 1$ or $x = \frac{H}{H-N}$. The marginal cost of a generating a unique address is

$$C'(N) = \frac{vH}{H-N} \quad (21)$$

and $C(N)$ is the indefinite integral of (21):

$$C(N) = vH(\ln(H) - \ln(H - N)). \quad (22)$$

Notice that $C'(N) > 0$ and $C''(N) > 0$.

Using this formulation, the stage 1 equilibrium mailing list N^* is

$$N^* = \frac{(\Pi(n^*) - v) H}{\Pi(n^*)}. \quad (23)$$

Equation (23) shows that when the cost of drawing an additional address is zero, the spammer targets the entire population ($N^* = H$), but when the cost of a draw is positive the spammer targets only a portion of the population ($N^* < H$).

Let the total number of messages sent by the spammer be

$$T = n^* N^*. \quad (24)$$

4.1 Comparative statics on N^*

The optimal size of the spammer's mailing list is decreasing with receiver pays pricing, sender pays pricing and the effectiveness of the filter because all these instruments decrease the spammer's expected profit per target:

$$\frac{\partial N^*}{\partial p^R} = \frac{\partial N^*}{\partial A} \frac{\partial A}{\partial \theta} \frac{\partial \theta(p^R)}{\partial p^R} = -\frac{vH}{\Pi(n^*)^2} \frac{\partial \Pi(n^*)}{\partial A} \frac{(c^{spam} + p^S)}{\theta^2 \alpha \pi} \frac{\partial \theta(p^R)}{\partial p^R} < 0, \quad (25)$$

$$\frac{\partial N^*}{\partial p^S} = \frac{\partial N^*}{\partial A} \frac{\partial A}{\partial p^S} = \frac{vH}{\Pi(n^*)^2} \frac{\partial \Pi(n^*)}{\partial A} \frac{\partial A}{\partial p^S} < 0 \quad (26)$$

and

$$\frac{\partial N^*}{\partial q} = \frac{vH}{\Pi(n^*)^2} \frac{\partial \Pi(n^*)}{\partial q} < 0. \quad (27)$$

In what follows, we use these comparative static results to discuss the impact of filtering and pricing on the total volume of spam.

4.2 Comparative statics on T

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

$$\frac{\partial T}{\partial p^R} = \frac{\partial n^*}{\partial p^R} N^* + \frac{\partial N^*}{\partial p^R} n^* < 0 \quad (28)$$

$$\frac{\partial T}{\partial p^S} = \frac{\partial n^*}{\partial p^S} N^* + \frac{\partial N^*}{\partial p^S} n^* < 0 \quad (29)$$

and

$$\frac{\partial T}{\partial q} = \frac{\partial n^*}{\partial q} N^* + \frac{\partial N^*}{\partial q} n^*. \quad (30)$$

Given (28) and (29) it is clear that increasing the receiver pays price and/or the sender pays price unambiguously reduce the total volume of spam. The impact of filtering on the total volume of spam in (30), 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 (27) we can see that the optimal mailing list decreases with q and so (30) 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 Further discussion

In this section we discuss three challenges that must be overcome when using pricing or filtering in the fight against spam; the problem of spammers and consumers taking evasive action to avoid paying prices and the problem of ISPs not willing to charge prices due to ISP competition.

5.1 ISP willingness to charge prices

We abstracted from any ISP behavior of any type when we modeled the decision of a spammer in Section 2. However, it seems reasonable to assume that competition between ISPs for consumers might lead some to charge lower (perhaps zero) receiver pays or sender pays prices than are necessary to eliminate spam.

One major issue that ISPs have to address when considering sender pays pricing is that individually they can do very little, if anything, to reduce the total amount of spam. An ISP charging a sender pays price on its own will simply drive their spammers to other networks (assuming here that relocation

costs for spammers are small relative to the benefits of avoiding sender pays prices) and so the total volume of spam will be unaffected. Of course the ISP's other consumers may choose to move as well (although the relocation costs for consumers might not be insignificant, if the benefits to ISPs of pricing are zero and the costs are small but positive, they won't fly) and so unilateral pricing will be unprofitable for an ISP. This introduces a coordination problem in that all ISPs benefit from the elimination of spam when all ISPs use sender pays pricing but, as long as there is one ISP that doesn't charge a price there might be little impact on spam volumes. This implies that there is little incentive for any ISP to introduce a sender pays price.

It is clear that there is a serious coordination problem but we believe that there is a simple technical solution to it. Imagine a world in which a number of ISPs coordinate on charging a sender pays price and agree to filter all messages sent from ISPs that do not charge a sender pays price. If this group was sufficiently large, consumers who did not want all of their messages sent to these ISPs to be filtered would quickly choose a conforming ISP and pay up. Nonconforming ISPs would then be left with only spammers in their potential customer base. However, spammers cannot be profitable if none of their messages get into consumers' inboxes so they too will choose never to subscribe with a nonconforming ISP. The ability to filter messages with 100 percent effectiveness on the basis of ISP origin means that, assuming that an initial critical mass of ISPs could be convinced to coordinate, there must be a stable equilibrium in which all ISPs charge a sender pays price.

A similar coordination problem exists in the introduction of receiver pays pricing. Here, as long as an ISP does not lose its entire customer base when it introduces a receiver pays price, it reduces the likelihood that a spammer's message reaches an interested customer and so reduces the expected marginal benefit of a spam message and, in turn, reduces the volume of spam. However,

this benefit is enjoyed by all email users and not just those who subscribe to the receiver pays price charging ISPs and so we have a classic free rider problem. While we do not need a complete uptake of receiver pays pricing to eliminate spam (we only need enough to drive the expected marginal benefit of a message below its marginal cost), full local coordination of ISPs will be needed to overcome the free-rider problem. This is because consumers will always prefer an ISP that does not charge for opening messages and therefore receiver pays pricing is sustainable in an equilibrium only if consumers have no incentive to switch. We can think of no good internal mechanism that will get enough regions to buy into the pricing scheme or to overcome the local free-rider problem.

5.2 Efforts to avoid paying prices

Our analysis in Section 2 suggests that sender pays and receiver pays pricing, used either alone or in conjunction with filtering, can be effective weapons against spam. For pricing to work in practice, however, one must first ensure that spammers and consumers are actually required to pay the price. Assuming that a mechanism exists whereby all emails that are sent and/or read are priced (we discuss whether this is likely below), the obvious problem that must be overcome is that along with prices comes an incentive to find ways to avoid paying them.

The magnitude of these spam eliminating prices (at fractions of a cent) will be small both absolutely and relative to the prices charged for other forms of communication. A consumer who sends and receives a small number of messages, therefore, has little incentive to undertake avoidance measures beyond being more discerning about what messages they choose to send and read. In the absence of practical substitutes for email, consumers are very likely to simply pay up⁶.

⁶Note that in New Zealand it costs upwards of ten cents to send a text message from a

For spammers, on the other hand, the incentive to avoid paying a spam eliminating sender pays price is significant because by its very construction, paying the price will render spam unprofitable. Spammers are already using illegal means to hack into consumers' computers, turning them into spammer zombies. This type of activity seems to be even more likely in the face of sender pays pricing because it would be the consumers whose computers were made into zombies who would pay for spam, not the spammer himself. However, this option might easily be closed to spammers as the risk of having to pay for spammers' messages should give consumers sufficient incentive to protect their computers from spammer attacks. One could also envisage a payment system for email messages similar to that for mobile telephones in which consumers prepay for a limited volume of activity. This implies that hacking would be a rather inefficient way for spammers to send a large number of email messages. Again, by removing the mechanism spammers have to circumvent pricing, this mechanism would help sender pays prices to have the desired effect to reduce spam.

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,

mobile phone yet millions of text messages are sent.

the more effective are receiver pays and sender-pays 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.

We highlight two practical issues associated with implementing prices. First, when consumers and spammers can choose an ISP amongst competing ISPs it becomes substantially more difficult for any one ISP to adopt pricing solutions in a war against spam. This implies that formal cooperation will be required for the implementation of these pricing strategies. On a positive note, we suggest a strategy of cooperation that does not require all ISPs to agree to cooperate initially. If a sufficient proportion of cooperating ISPs agree to set spam-eliminating sender pays prices as well as to filter all messages coming from non-cooperating ISPs’ customers, then we might be able to achieve full cooperation that is sustainable.

We also discuss the practical problem with implementing pricing that arises when spammers can take measures to avoid having to pay the prices, rendering pricing ineffective. Spammers already hack into consumers’ computers and make them into zombies that send bulk email from the spammer in order to prevent their ISPs from detecting them as spammers. The incentive to engage in this type of activity would certainly go up if spammers had to pay for sent messages. The victims of these attacks would have to pay for the messages sent from their computers, reducing the ability of sender pays pricing to reduce spam. However, we envision that the prospect of having to pay for millions of messages would give consumers the right incentives to protect their computers from hacker attacks. Furthermore, other systems whereby only a small number of messages are pre-paid for could be implemented to reduce the benefit to spammers from hacking

into consumers' computers.

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] Howell, D. (2004) "E-mail spammers need only a few customers to get meaty rewards." *Investor's Business Daily*, 19 February 2004.
- [6] Khong, D. (2004) "An economic analysis of spam laws", *Erasmus Law and Economics Review* 1, 23-45.
- [7] 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.
- [8] Loder, T., Van Alstyne, M. and R. Wash (2006) "An Economic Response to Unsolicited Communication," *Advances in Economic Analysis & Policy* 6(1), Article 2.
- [9] The Economist (2007) "Spam seems here to stay." *Seattle Post - Intelligencer*, September 28, 2007.

- [10] 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.
- [11] Van Alstyne, M. (2007) “Curing spam: rights, signals & screens.” *Economists’ Voice*, 4(2), March 2007.