

# The Effects of Anti-Spam Methods on Spam Mail

Eran Reshef  
Blue Security Inc., [www.bluesecurity.com](http://www.bluesecurity.com)

Eilon Solan  
School of Mathematical Sciences  
Tel Aviv University  
Tel Aviv, 69978 Israel  
[eilons@post.tau.ac.il](mailto:eilons@post.tau.ac.il)

## ABSTRACT

We provide a model to study the effects of three methods of fighting spam mail, namely (1) increasing the cost of mailing messages, (2) filters, and (3) a do-not-spam registry, on the number of spam messages that users receive, and on the efficiency of the internet (measured by the total number of spam messages spammers send).

## 1. INTRODUCTION

Bulk electronic mail, also known as spam mail, has become a major danger to the efficiency of the internet. Postini, a provider of email security,<sup>1</sup> reported in May 2006 that spam activity has increased over 65% since January, 2002, and that more than 80% of e-mail transportation is spam. The report states that “this increase causes e-mail systems to experience unexpected overload in bandwidth, server storage capacity, and loss of end-user productivity.”

There are several approaches that have been discussed in the literature to defend oneself from spam mail. The most popular approach is to use a filter, which is supposed to filter out spam messages. The filter does indeed reduce the number of spam messages that the user receives, but it does not entirely eliminate the problem. Moreover, filters have inherent problems; for example, they are difficult to maintain, senders of spam mail adapt to their strategies, and they sometimes filter innocent mail. The effectiveness of this method has been discussed in numerous articles, including, e.g., [3], [4], [5], [8], [9], [13].

Another approach is to increase the cost of mailing spam messages. This goal can be achieved in various ways (see, e.g., [8], [21]): authentication and reputation services, counter attacks, channelling (e.g., [15], [12]), payments, either monetary or in computational power (e.g. [1], [10], [17], [18], [20], [23], [30]), digital signatures (e.g., [29]), and regulatory actions (e.g., [16]).

A third approach is a do-not-spam registry (or any tech-

<sup>1</sup>See [www.postini.com](http://www.postini.com).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CEAS 2006 - *Third Conference on Email and Anti-Spam*, July 27-28, 2006, Mountain View, California USA  
Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

nology that ensures its users do not receive spam messages). In this solution, users who opt not to receive spam mail join a do-not-spam registry, and spammers should not mail those users any spam mail. The American Federal Trade Commission has studied the feasibility of such a registry, and concluded in June 2004 that “a National Do Not Email Registry, without a system in place to authenticate the origin of email messages, would fail to reduce the burden of spam and may even increase the amount of spam received by consumers.” Nevertheless, attempts at launching a registry have been made (e.g., the Michigan and Utah registry for children in summer 2005). A commercial attempt by Blue Security inc. to initiate a registry that involves sending an opt-out if a spam mail is received was partially successful, as 6 out of the top 10 spammers complied with the registry. However, certain spammers viewed this solution as a strategic threat to their spam business. Those spammers waged a full scale cyber war on Blue Security and its users. Blue Security was forced to cease its anti-spam operation in order to prevent an escalation that would have brought down the Internet.<sup>2</sup>

Current technology’s ability to support any of these methods is a crucial step towards successful implementation. However, another important question that hasn’t been asked is what are the effects of each of the methods on spam transportation. That is, suppose better filters are introduced. How will that affect the number of spam messages spammers send (affecting the efficiency of the internet)? Or, suppose there is an effective do-not-spam registry, and suppose that some portion of the population joins it. Will other users, who do not join the registry, receive more or less spam messages?

In the present paper we present a game theoretic model that enables us to study the effects of the three approaches mentioned above on the amount of spam messages spammers send, and on the amount of messages users receive.

The model involves users, spammers, and spam messages that spammers send to users. We take the point of view of a social planner, whose goal is to improve the efficiency of the internet, or of the majority of the users, who wish to reduce the amount of spam mail they receive. Therefore, the goal is to reduce the number of spam messages that spammers send or users receive, and to this end we study the effects of the three approaches on spam transportation.

We study in detail the optimization problem of the spammers: what is the optimal number of messages the spammers should send to the users. As there are many spam-

<sup>2</sup>see [28] for more information.

mers, the optimization problem is multi-dimensional, or a game. Often in optimization problems, especially when several decision makers are involved, there are several optimal solutions, and therefore a question arises concerning which solution will be chosen by society. We prove that in *all* optimal solutions the total number of messages that are sent to each user is the *same*, so that, from the point of view of the users, or that of a social planner, it does not matter which optimal solution is chosen by the spammers (Theorem 5).

We go on and study qualitative properties of the solution. We show that increasing the cost of sending spam messages has the unequivocal effect of reducing the total number of spam messages sent by spammers, and therefore also the amount of spam messages users receive (Theorem 6). This result has been discussed in the literature (see, e.g., [18]), but to the best of our knowledge, our model is the first in which the result is formally proven. (See [30] for the opposite conclusion when spammers have accurate information on the users).

On the other hand, improving the quality of the filter has an ambiguous effect: the number of messages received by the user whose filter is improved indeed decreases (Theorem 7), but the total number of messages sent by spammers might increase (Theorem 8). This implies that (1) other users, whose filters are not improved, might receive more spam messages, and (2) the internet might become less efficient. It turns out that this ambiguous effect happens when the cost of sending spam messages is not too high. When this cost happens to be high, the effect of improving filters is unequivocally positive. We provide a simple condition that indicates the minimal cost above which improving filters is beneficial for the whole society. An interesting implication of our results is that, as long as the cost of sending spam is not high, social efficiency (measured by the total number of spam messages sent by spammers) requires inefficient filters; once filters reach a certain quality, any additional improvement makes the internet more crowded.

Regarding a registry, we show that when the high-quality filters operate in the range in which its improvement is inefficient to society, the registry has the following effect: if users of a low-quality filter join the registry, the total number of spam messages sent by spammers increases, whereas if users of a high-quality filter join the registry, the total number of spam messages sent by spammers decreases (Theorem 11). We go on to show that as more users of low-quality filters join the registry, the total number of spam messages sent to users decreases, so that having these users join the registry improves the internet's efficiency (Theorem 13).

The main conclusion learned from our results is that to effectively fight spam messages the cost of sending spam messages must increase. Without such a measure, the huge investment in improving filters only causes flooding the internet with more spam messages, thereby reducing its efficiency, and flooding users who do not have the state-of-the-art filter with more and more spam messages.

Another conclusion concerns pricing the use of filters and that of a registry. Since their use may increase the number of spam messages spammers send, the users who are potentially most harmed by these technologies are poor users, who cannot afford employing these technologies. If high-quality filters and/or registering is costly, poor users will suffer from a technology that helps rich users. Therefore, the socially responsible solution is to offer these technolo-

gies free of charge to poor users. Furthermore, the registry serves as an improved substitute to users of high-quality filters: it completely defends its users from spam mail while improving the internet's efficiency.

A parallel to our model in consumer theory is a market with several consumers and two goods (here the consumers are the spammers, and the goods are the two populations of users). Though some spammers target their messages to a specific group of users, most spammers do not distinguish between different users. In the language of the model from consumer theory, the goods are complementary, and the consumer is restricted to consume the same amount of each good. Unlike standard consumer theory, here prices (quality of the filters) are exogenous and consumers react to the average price.

The closest papers to ours are [3], [9], [16], and [30], which provide game theoretic analysis of certain games between spammers and users. [3] analyzes a game between a single spammer and a single user, in which the spammer has to decide how many spam messages to send, and the user can tune his or her filter to filter out more or less messages (bad and honest). They find the optimal level of filtering from the users' point of view. [30] studies the equilibrium in a game played between senders and receivers, where senders can target receivers, receivers are characterized by their type, which indicates the messages that interest them, and they are bounded by the number of messages they can process. They show that if the cost of sending messages is sufficiently low, senders sales are increasing in the cost, and, if senders have enough information on receivers characteristics, then senders profits are also increasing in the cost. [16] provides an economic study of the spam problem with an emphasis on the legal aspect. Our work can be viewed as an extension of [16] with an emphasis on technological solutions. [9] studies a classification problem as a game between a classifier and a single adversary, in which each player adapts to the strategy the other employs as the game continues.

Our model differs from these papers in several respects. Whereas [3] takes the filtering technology as given, assumes that all users have the same technology, and the decision problem of users is how to tune their filter, we assume that different users have different filters which may improve over time, and we take the users as static: the decisions regarding improving the filters are exogenous (they are taken by the internet provider or by the developer of the filter). [30] assumes that users do not read messages only because of exogenous constraints, and that senders have information on the receivers and can send targeted messages. We, on the other hand, assume that users are not interested in spam messages at all, and that senders have no information on the receivers. Finally, [16] does not study the effect of technological solutions, and studies the problem only from the point of view of social efficiency.

The paper is organized as follows. The formal model is presented in section 2. In section 3 we discuss the stable states that the market will reach, and in section 4 we study in detail the effects of changing the environment on the amount of spam messages that spammers send and users receive. In section 5 we discuss which assumptions can be weakened, and possible extensions of the model that are left for future research.

## 2. THE MODEL

We now present the formal model that we study.

## 2.1 The population

There are  $M$  users, each of whom uses a filter to filter out spam messages. The quality of a filter is measured by the percentage of spam messages it lets through (the *waste*). The users are divided into two sub-populations, according to the quality of the filters they use.<sup>3</sup> We denote by  $w_1$  and  $w_2$  the quality of the filters of the two sub-populations. The case  $w_2 = 1$  is equivalent to having a group of users who do not use any filter. We call users who use a filter with quality  $w_1$  (resp.  $w_2$ ) *type-1 users* (resp. *type-2 users*). We denote by  $M_1$  the number of type-1 users, and by  $M_2$  the number of type-2 users, so that  $M_1 + M_2 = M$ . Our analysis ignores the the percentage of false-positives – the percentage of legitimate messages that are classified as spam. This is so because we take the users as static – they can neither tune their filters nor switch to a different filter. The only active participants in our model are the spammers, who only care about the percentage of false-negatives.

Though in our model it is undesirable to receive spam messages, there are users, termed *potential buyers*, who may purchase a product that is advertised in spam messages (a *spam product*). We denote the percentage of potential buyers among all users by  $\mu$ . We assume that potential buyers are evenly distributed among the population.<sup>4</sup> That is, the percentage of potential buyers among both type-1 users and type-2 users is  $\mu$ . To simplify the analysis, we suppose that each potential buyer may spend a fixed amount of  $T$  dollars on spam products every year, and he or she does so in a single purchase. Alternatively,  $T$  may be thought of as the average annual spending of a user on spam products.

We denote the probability that a potential buyer will not purchase the product that is advertised in any given spam message by  $p_*$ . That is, when a potential buyer receives a spam message, with probability  $p_*$  he or she deletes it, and with probability  $1 - p_*$  he or she spends  $T$  dollars on the spam product advertised in the message. Once this user spends his or her  $T$  dollars, he or she will not purchase any other spam product that year.

The users in our model are static: they make no decisions at all – neither to switch to a different filter, nor to quit using e-mails. The second assumption is valid as in practice users keep on using their e-mail even though they do receive spam messages; the constraint that the user's cost of handling spam mail is lower than his or her gain from using e-mail is not binding, and hence can be ignored. The first assumption is indeed restricting, but we believe it can be made as in practice most users do not change their internet provider. Another interpretation of the assumptions is that the model analyzes the behavior of *new* users who join the internet. Once a user gains experience he or she will not purchase a spam product in any case, and is effectively taken out of the spammers' optimization problem.

## 2.2 The spammers

<sup>3</sup>The model can be extended to more than two sub-populations at the cost of additional complexity.

<sup>4</sup>Recent surveys find that roughly 20% of the users say that they purchased spam products (see Yahoo Mail Survey - <http://www.cbsnews.com/stories/2004/06/29/tech/main626625.shtml> or Forrester Research - <http://news.zdnet.com/2100-1009.22-5487375.html>).

There are  $N$  spammers. The decision problem of each spammer is how many spam messages to send to each user each year. We assume that spammers do not distinguish between users, so that a spammer mails the same number of messages to all users.<sup>5</sup> We also assume that the users do not distinguish between the spammers, so that the percentage of potential buyers who purchase from any given spammer is given by the percentage of spam messages that user receive from the spammer. That is, if a given spammer sends  $x$  messages to a given user, while all other spammers send  $y$  messages to that user, then, provided the user purchases a spam product, the probability he or she will purchase it from that spammer is  $x/(x + y)$ .

We now describe the payoff function of a spammer. Denote by  $x$  the number of messages per user per year that a specific spammer sends, and by  $y$  the total number of spam messages per user per year sent by all other spammers. Denote the cost per message by  $d$  dollars. The payoff to the spammer is<sup>6</sup>

$$W(x, y) = -xMd + \frac{x}{x+y}T\mu \left( M_1 \left( 1 - p_*^{w_1(x+y)} \right) + M_2 \left( 1 - p_*^{w_2(x+y)} \right) \right). \quad (1)$$

Indeed,  $xMd$  is the total cost of sending  $x$  messages to all users,  $1 - p_*^{w_k(x+y)}$  is the probability that a type- $k$  user who receives  $x + y$  messages will purchase a spam product, and  $\frac{x}{x+y}$  is the percentage of users who purchase from the spammer among all users who purchase a spam product.

The goal of each spammer is to maximize his or her expected gain. If  $W(x, y)$  is lower than the spammer's fixed cost the spammer will go out of the market. Otherwise, the spammer makes a profit.

We note that spammers do know the quality of the filters users use: first, these figures can be found on the internet, and second, in practice spammers purchase filters, learn their technology, and find ways to circumvent them.

Finally, the fact that there is more than one spammer drives spam volume up, since a spammer who sends more spam messages gains in two ways: first, the market increases, as more users will purchase spam products, and second, since the ratio  $x/(x + y)$  increases with  $x$ , the spammer's market share increases as well. When there is a single spammer, only the first effect occurs.

## 2.3 Stable configurations

For the mathematical analysis it is more convenient to assume that  $x$ , the number of spam messages per user per year the spammer sends, need not be a non-negative integer, but can be any non-negative real number.

A vector  $\vec{x} = (x_1, \dots, x_N)$ , which indicates the number of spam messages each spammer sends, is termed a spam configuration.

DEFINITION 1. A spam configuration (or simply a configuration) is a vector of non-negative real numbers  $\vec{x} =$

<sup>5</sup>Thus, targeting a sub-population is not allowed. In practice, most spammers do not target sub-populations. In case spammers use a targeting strategy, the amount of spam messages is lower, see, e.g., [18] and [30].

<sup>6</sup>The fixed cost of the spammer is not included in the spammer's payoff as it has no effect on the analysis that follows.

$(x_1, \dots, x_N)$ , where for each  $i$  the quantity  $x_i$  is the number of spam messages per user per year sent by spammer  $i$ .

The market reaches a stable state if no spammer finds it profitable to change the number of spam messages he mails to each user. Usually, decision makers adapt to small changes in the environment by slightly changing their behavior, and they do not consider significant changes that may increase their profit. We therefore define a spam configuration to be stable if no spammer can profit by slightly changing the number of spam messages he or she sends.

DEFINITION 2. A spam configuration  $\vec{x} = (x_1, \dots, x_N)$  is stable if, for every spammer  $i$ ,<sup>7</sup>

$$\frac{\partial W}{\partial x} \left( x_i, \sum_{j \neq i} x_j \right) = 0.$$

Observe that a stable spam configuration is not necessarily a Nash equilibrium, as spammers may profit by large changes in their behavior. Rather, it is an inflection point of the function  $V : \mathbf{R}_+^N \rightarrow \mathbf{R}^N$ , where  $V_i(\vec{x}) = W \left( x_i, \sum_{j \neq i} x_j \right)$ .

## 2.4 Direct cost and direct gain

The *direct cost* of mailing a single spam message is  $d$ . We define the *direct gain* from a single spam message by

$$g = T\mu(-\ln p_*)(M_1 w_1 + M_2 w_2)/M. \quad (2)$$

We now explain the motivation of this definition. Consider the first message that is sent. With probability  $\mu$  the user who receives it is a potential buyer, with probability  $(M_1/M)w_1 + (M_2/M)w_2$  the message is not filtered, and with probability  $1 - p_*$  it causes the user to purchase a spam product. Therefore, the first message sent to a random user has probability  $\mu(1 - p_*)(M_1 w_1 + M_2 w_2)/M$  of making that user purchase a spam product. In practice,  $p_*$  is close to 1, and therefore  $(-\ln p_*)$  is close to  $1 - p_*$ . As the amount spent in purchasing a spam product is  $T$  dollars,  $g$  represents the expected income from the first message that is sent to a random user.<sup>8</sup>

As we show below, the behavior of the market is different when  $g > d$  and when  $g \leq d$ .

## 3. EXISTENCE AND UNIQUENESS OF A SOLUTION

The first result states that if the direct gain is lower than the direct cost, there will be no spam.

THEOREM 3. Suppose that  $g \leq d$ , and let  $\vec{x} = (x_1, \dots, x_N)$  be any configuration. If  $\sum_{i=1}^N x_i > 0$  then there is  $i$  such that (1)  $x_i > 0$ , and (2)  $W \left( x_i, \sum_{j \neq i} x_j \right) < 0$ .

Thus, if the direct gain is lower than the direct cost, and some spammers do mail spam messages, at least one of those spammers suffers a loss, and will be driven out of the market. Iterating this argument shows that there will be no spam.

<sup>7</sup>Recall that  $W(x, y)$  is a function of two variables, so that  $\partial W/\partial x$  is its derivative relative to its first argument.

<sup>8</sup>Our proof below holds also when  $p_*$  is not close to 1. We here provide only the intuition.

Before proving the theorem let us calculate the derivative of  $W$ , which we need in the sequel:

$$\begin{aligned} \frac{\partial W}{\partial x}(x, y) &= -Md \\ &+ \frac{x}{x+y} T\mu(-\ln p_*) \left( M_1 w_1 p_*^{w_1(x+y)} + M_2 w_2 p_*^{w_2(x+y)} \right) \\ &+ \frac{y}{(x+y)^2} T\mu \left( M_1 (1 - p_*^{w_1(x+y)}) + M_2 (1 - p_*^{w_2(x+y)}) \right). \end{aligned} \quad (3)$$

PROOF. Assume w.l.o.g. that  $x_i > 0$  for every  $i$  (otherwise, drop from the market the spammers who do not send spam messages). Denote  $z = \sum_{i=1}^N x_i$ , and observe that<sup>9</sup>

$$\begin{aligned} W(x_i, z - x_i) &= -x_i M d \\ &+ \frac{x_i}{z} T\mu \left( M_1 (1 - p_*^{w_1 z}) + M_2 (1 - p_*^{w_2 z}) \right). \end{aligned} \quad (4)$$

Summing (4) over  $i = 1, \dots, N$ , and using Lemma 14 (see Appendix) we obtain

$$\begin{aligned} \sum_{i=1}^N W(x_i, z - x_i) &= \\ &= -z M d + T\mu \left( M_1 (1 - p_*^{w_1 z}) + M_2 (1 - p_*^{w_2 z}) \right) \\ &< z \left( -d M + \frac{T\mu}{z} \left( M_1 (1 - p_*^{w_1 z}) + M_2 (1 - p_*^{w_2 z}) \right) \right) \\ &< z (-d M + T\mu (M_1 (-\ln p_*) w_1 + M_2 (-\ln p_*) w_2)) \\ &= z M (g - d). \end{aligned}$$

Hence, if  $g \leq d$ , we have  $\sum_{i=1}^N W(x_i, z - x_i) < 0$ , so that at least one spammer receives a negative payoff.  $\square$

We now show that a stable configuration exists as soon as the direct gain exceeds the direct cost.

THEOREM 4. If  $g > d$ , a stable configuration exists.

PROOF. We show that there is a stable configuration in which all spammers mail the same number of spam messages. Substituting  $(x, y) = \left( \frac{1}{N}z, \frac{N-1}{N}z \right)$  in (3), we obtain

$$\begin{aligned} \frac{\partial W}{\partial x} \left( \frac{1}{N}z, \frac{N-1}{N}z \right) &= \\ &= -Md + (-\ln p_*) \frac{T\mu}{N} \left( M_1 w_1 p_*^{w_1 z} + M_2 w_2 p_*^{w_2 z} \right) \\ &+ \frac{(N-1)T\mu}{Nz} \left( M_1 (1 - p_*^{w_1 z}) + M_2 (1 - p_*^{w_2 z}) \right). \end{aligned}$$

One can easily verify that

$$\lim_{z \rightarrow \infty} \frac{\partial W}{\partial x} \left( \frac{1}{N}z, \frac{N-1}{N}z \right) = -Md < 0. \quad (5)$$

Using (26) (see Appendix) we obtain

$$\lim_{z \rightarrow 0} \frac{\partial W}{\partial x} \left( \frac{1}{N}z, \frac{N-1}{N}z \right) = M(g - d). \quad (6)$$

Since  $g > d$ , the limit in (6) is positive. Since the function  $z \mapsto \frac{\partial W}{\partial x} \left( \frac{1}{N}z, \frac{N-1}{N}z \right)$  is continuous, it follows from (5) and (6) that there is  $z_* > 0$  such that

$$\frac{\partial W}{\partial x} \left( \frac{1}{N}z_*, \frac{N-1}{N}z_* \right) = 0. \quad (7)$$

Define the configuration  $\vec{x} = (z_*/N, \dots, z_*/N)$ . By (7)  $\vec{x}$  is stable.  $\square$

<sup>9</sup>Throughout,  $z$  denotes the total number of spam messages sent to each user.

In many models where stable points exist, there exist many such stable points, and then a question arises concerning which stable point the market will reach, and which factors, that were not taken into account in the analysis, influence the outcome. The next result states that from the point of view of the users, all stable configurations are equivalent – each user receives the same number of spam messages in all stable configurations. Thus, even though there might be several stable configurations, in which the total amount of spam is divided in different *ways* among the spammers, the users are indifferent among those stable configurations.

**THEOREM 5.** *Let  $\vec{x} = (x_1, \dots, x_N)$  and  $\vec{x}' = (x'_1, \dots, x'_N)$  be two stable configurations. Then  $\sum_{i=1}^N x_i = \sum_{i=1}^N x'_i$ .*

The intuition behind this result is as follows. The gain from sending one additional spam message depends on the total number of spam messages all spammers send, and not on the exact number of messages each spammer sends. Moreover, it decreases with this quantity. In a stable configuration, this gain is equal to the cost of sending that message. Since the cost of sending a single spam message is constant, in any stable configuration the total number of spam messages that are sent is constant as well.

The gain from sending one additional spam message, as well as the cost of sending it, are the same for all spammers, since spammers do not target users. This might suggest that Theorem 5 is particular to the model we study. However, the theorem is true in a much more general setup: suppose the users are divided into sub-populations, and each spammer targets several sub-populations. Then as long as the characteristics of the sub-populations are the same (the same percentage of type-1 and type-2 users, and the same percentage of potential buyers), then all users will receive the same total number of spam messages in all stable configurations.

Recall that  $M = M_1 + M_2$ . Define

$$\begin{aligned} F &= F(w_1, w_2, M_1, M_2, z) \\ &= (-\ln p_*) \left( \frac{M_1}{M} w_1 p_*^{w_1 z} + \frac{M_2}{M} w_2 p_*^{w_2 z} \right) \\ &\quad + \frac{N-1}{z} \left( \frac{M_1}{M} (1 - p_*^{w_1 z}) + \frac{M_2}{M} (1 - p_*^{w_2 z}) \right). \end{aligned}$$

As we will see below, this function is related to the derivative of  $W$ .

By Lemma 15 (see Appendix), and since the function  $z \mapsto p^z$  is monotonic decreasing, we obtain that  $\partial F / \partial z < 0$ .

We are now ready to prove Theorem 5.

**PROOF.** Let  $\vec{x} = (x_1, \dots, x_N)$  be a stable configuration, and denote  $z = \sum_{i=1}^N x_i$ . Then

$$\begin{aligned} 0 &= \sum_{i=1}^N \frac{\partial W}{\partial x} (x_i, z - x_i) \\ &= -NMd + T\mu(-\ln p_*) (M_1 w_1 p_*^{w_1 z} + M_2 w_2 p_*^{w_2 z}) \\ &\quad + \frac{N-1}{z} T\mu (M_1 (1 - p_*^{w_1 z}) + M_2 (1 - p_*^{w_2 z})) \\ &= MT\mu \left( -\frac{dN}{T\mu} + F \right). \end{aligned}$$

In particular,  $z$ , the total number of spam messages sent to

each user, is a solution of

$$F = F(w_1, w_2, M_1, M_2, z) = \frac{dN}{T\mu}. \quad (8)$$

Since  $F$  is monotonic decreasing in  $z$ , there can be at most one solution  $z$  to (8), and the result follows.  $\square$

## 4. QUALITATIVE PROPERTIES OF THE SOLUTION

After we established the general existence of a stable configuration, and its uniqueness from the point of view of the users, we turn to study qualitative properties of stable configurations. Namely, how changes in the environment affect the total number of messages that are sent by spammers or received by users.

For every  $w_1, w_2, M_1, M_2$ , and  $d$ , denote the total number of messages per user per year sent by spammers in all stable configurations by  $z = z(w_1, w_2, M_1, M_2, d)$ . By Theorem 5,  $z$  is well defined. We will now study how  $z$  depends on its parameters.

### 4.1 Effects of Increasing Cost

Intuition suggests that as the cost of sending mails increases, the amount of spam messages decreases. As the following theorem states, this is indeed the case.

**THEOREM 6.** *For every fixed  $w_1, w_2, M_1$ , and  $M_2$  the function  $d \mapsto z(w_1, w_2, M_1, M_2, d)$  is monotonic decreasing.*

The intuition is the same as for Theorem 5: the marginal gain from each additional spam message decreases with the total number of messages already sent. Since in a stable configuration the marginal gain is equal to the cost of sending that message, increasing the cost must imply a decrease in the total number of messages sent in a stable configuration.

**PROOF.** By (8) we have  $F \circ z = dN/T\mu$ , so that by the chain rule<sup>10</sup>

$$0 < \frac{N}{T\mu} = \frac{\partial(F \circ z)}{\partial d} = \frac{\partial F}{\partial z} \frac{\partial z}{\partial d}.$$

Since  $\partial F / \partial z < 0$  we deduce that  $\partial z / \partial d < 0$ , and this is the desired result.  $\square$

### 4.2 Effects of Improving Filters

In the present section we study in detail the effects of improving the filter of one of the two sub-populations on the number of messages spammers send or users receive.

One can verify that for every fixed  $M_1, M_2$ , and  $d$ , the function  $z$  is differentiable at  $(w_1, w_2, M_1, M_2, d)$  w.r.t.  $w_1$  and  $w_2$  for almost every  $w_1$  and  $w_2$ . Therefore studying the effects of the parameters of  $z$  on its value reduces to the study of its directional derivatives.

The following theorem states that a user gains when the filter he or she uses improves.

**THEOREM 7.** *Fix  $p_*, w_2, M_1, M_2$ , and  $d$ . The function  $w_1 \mapsto w_1 \cdot z(p_*, w_1, w_2, M_1, M_2, d)$  is monotonic increasing.*

Plainly an analog statement holds for type-2 users.

<sup>10</sup>To simplify writing, we shorten  $F(w_1, w_2, M_1, M_2, z(w_1, w_2, M_1, M_2, d))$  to  $F \circ z$ .

PROOF. Set  $u(w_1, w_2, M_1, M_2, d) = w_1 z(p, w_1, w_2, M_1, M_2, d)$ . *if and only if*  
This is the total number of spam messages that bypass the filter of a type-1 user. Set

$$H(w_1, w_2, M_1, M_2, u) = F \left( w_1, w_2, M_1, M_2, \frac{u}{w_1} \right).$$

Then,

$$\begin{aligned} M \times H(w_1, w_2, M_1, M_2, u) = & \\ & (-\ln p_*) (M_1 w_1 p_*^u + M_2 w_2 p_*^{w_2 u / w_1}) \\ & + \frac{w_1(N-1)}{u} (M_1(1-p_*^u) + M_2(1-p_*^{w_2 u / w_1})). \end{aligned}$$

Since the function  $u \mapsto p^{cu}$  is monotonic decreasing for every  $c > 0$ , Lemma 15 (see Appendix) implies that  $\partial H / \partial u < 0$ . Since the function  $x \mapsto x(1-p^{a/x})$  is monotonic increasing for every  $p \in (0, 1)$  and every  $a > 0$  on the range  $(0, \infty)$ , we obtain that  $\partial H / \partial w_1 > 0$ .

Observe that

$$H(w_1, w_2, M_1, M_2, w_1 z(w_1, w_2, M_1, M_2, d)) = \frac{dN}{T\mu}.$$

We deduce that the derivative of the left-hand side w.r.t.  $w_1$  vanishes. By the chain rule this derivative is equal to

$$0 = \frac{\partial H}{\partial w_1} + \frac{\partial H}{\partial u} \frac{\partial u}{\partial w_1}.$$

Since  $\partial H / \partial w_1 > 0$  while  $\partial H / \partial u < 0$  it follows that  $\partial u / \partial w_1 > 0$ , which is the desired result.  $\square$

Though improving the filter is beneficial for the user who uses it, as the following result shows this is not the case for other users.

We explore this issue now.

**THEOREM 8.** *The function  $w_1 \mapsto z(p_*, w_1, w_2, M_1, M_2, d)$  is monotonic decreasing if and only if*

$$\begin{aligned} & M_1 w_1 \left( e^{-N} + \frac{N-1}{N} (1 - e^{-N}) \right) \\ & + M_2 \left( w_2 e^{-w_2 N / w_1} + \frac{N-1}{N} w_1 (1 - e^{-w_2 N / w_1}) \right) \\ & > \frac{dMN}{T\mu(-\ln p_*)}, \end{aligned} \quad (9)$$

*and monotonic increasing if the opposite inequality holds.*

Thus, suppose that the function  $w_1 \mapsto z(p_*, w_1, w_2, M_1, M_2, d)$  is monotonic decreasing. This means that as  $w_1$  decreases (i.e., filters of type-1 users improve)  $z$  increases (i.e., more spam messages are sent). Since the right-hand side in (9) depends linearly on  $d$ , this result implies that if the direct cost of sending a single message is not too high, improving the quality of the filters of type-1 users has the undesirable effect of increasing the total spam messages sent by spammers. An analog statement for type-2 users holds as well.

The proof of Theorem 8 is similar to that of Theorem 7, and hence omitted.

In practice  $N$  is large, so that  $e^{-N} \approx 0$ . Since  $w_2 > w_1$ ,  $\exp(-w_2 N / w_1) \approx 0$  as well. Therefore Eq. (9) can be simplified as follows.

**COROLLARY 9.** *Suppose  $w_1 < w_2$ . When  $N$  is large, the function  $w_1 \mapsto z(w_1, w_2, M_1, M_2, d)$  is monotonic decreasing*

$$w_1 > \frac{dN^2}{(N-1)T\mu(-\ln p_*)}, \quad (10)$$

*and monotonic increasing if the opposite inequality holds.*

Eq. (10) provides a simple practical criterion for determining whether improving the high-quality filter has an unequivocal effect on spam mail, or whether it hurts the internet's efficiency and the welfare of the users who use the low-quality filter.

Our estimates are that  $N \approx 200$ ,  $d \approx \$0.000001$ , and  $T \approx \$100$ . Substituting  $\mu \approx 20\%$  and  $p_* \approx 99\%$  yields that improving filters increases efficiency if and only if  $w_1 < 0.1\%$ . As presently filters operate at a quality  $w_1 \approx 4\%$ , improving filters makes the internet more crowded, and hurts users who do not use filters.

Corollary 9 implies that as long as the high-quality filters are not sufficiently good, that is, as long as  $w_1 > dN^2 / (N-1)T\mu(-\ln p_*)$ , internet's efficiency requires a decrease in the quality of the high-quality filters.

### 4.3 Effects of a Registry

We now turn to study the effects of a registry on spam transportation. Since we assume that spammers adhere to the registry, a user who joins the registry is effectively removed from the population. Hence we study the effect of changing  $M_1$  and  $M_2$  on  $z$ , the total number of messages spammers send each non-registered user.

One can verify that if  $w_1 \neq w_2$  then for every fixed  $w_1, w_2$ , and  $d$ , the function  $z$  is differentiable at  $(w_1, w_2, M_1, M_2, d)$  w.r.t.  $M_1$  and  $M_2$  for almost every  $M_1$  and  $M_2$ .

Our first qualitative result concerning a registry is that the effect of type-1 users registering on the total number of spam messages sent is *always* opposite to the analog effect when type-2 users join the registry. That is, if having a type-1 user join the registry decreases the number of spam messages spammers send, then having a type-2 user register increases the number of spam messages spammers send. On the other hand, if having a type-1 user join the registry increases the number of spam messages spammers send, then having a type-2 user join decreases the number of spam messages spammers send.

This result is quite surprising, as it says that there is no win-win situation: in any given situation, either society as a whole prefers that type-2 users join the registry, or it prefers that type-1 users do so, but not both. The finding is formally stated by the following theorem.

**THEOREM 10.** *If  $w_1 \neq w_2$  then*

$$\frac{\partial z}{\partial M_1} \frac{\partial z}{\partial M_2} < 0.$$

The main force behind this result is that the behavior of spammers depends on  $M_1$  and  $M_2$  only through the ratio  $M_1/M_2$ . When type-1 users join the registry  $M_1$  decreases and therefore the ratio decreases as well, whereas when type-2 users join  $M_2$  decreases and therefore the ratio increases. Thus, the effect of having type-1 users join the registry on the ratio is opposite to the effect of having type-2 users join, so that the effect on the behavior of spammers is opposite as well.

It is interesting to note that this result is not particular to the model we study, but it is valid in any model in which the

behavior of spammers depends on  $M_1$  and  $M_2$  only through the ratio  $M_1/M_2$ .

PROOF. Define  $\beta = \frac{M_1}{M_1+M_2}$ ; it is the percentage of type-1 users in the population.  $F$  depends on  $M_1$  and  $M_2$  only through the ratio  $\beta = M_1/(M_1 + M_2)$ . Present  $F$  as a function of  $w_1, w_2, \beta$  and  $z$ . By (8) and by the chain rule,

$$0 = \frac{\partial(F \circ z)}{\partial\beta} = \frac{\partial F}{\partial\beta} + \frac{\partial F}{\partial z} \frac{\partial z}{\partial\beta}.$$

Since  $\partial F/\partial z < 0$ , the sign of  $\partial z/\partial\beta$  is the same as the sign of  $\partial F/\partial\beta$ . Now,  $F$  is linear in  $\beta$ , with a coefficient

$$(-\ln p_*) (w_1 p_*^{w_1 z} - w_2 p_*^{w_2 z}) + \frac{N-1}{z} (p_*^{w_2 z} - p_*^{w_1 z}). \quad (11)$$

Therefore  $\partial z/\partial\beta$  is positive if and only if the quantity in (11) is positive. Now, since  $\beta$  is increasing in  $M_1$  and decreasing in  $M_2$ ,  $\partial z/\partial\beta$  is positive if and only if  $\partial z/\partial M_1$  is positive, and if and only if  $\partial z/\partial M_2$  is negative. It follows that  $\partial Z/\partial M_1$  is positive if and only if  $\partial Z/\partial M_2$  is negative, and the result follows.  $\square$

The following theorem states that when filters operate in the range where improvements in the high-quality filters do not have a positive impact on the society as a whole, then having users with high-quality filters join the registry is beneficial to society. Therefore, in this case by Theorem 10 having users with low-quality filters join a registry harms the internet's efficiency.

**THEOREM 11.** *Assume that  $w_1 < w_2$ . Provided  $N$  is large, if  $w_1 > \frac{dN^2}{(N-1)T\mu(-\ln p_*)}$  then  $\frac{\partial z}{\partial M_1} > 0$ , so that by Theorem 10  $\frac{\partial z}{\partial M_2} < 0$ .*

PROOF. As we saw in the proof of Theorem 10,  $\partial Z/\partial M_1$  is positive if and only if the coefficient in (11) is positive. Set  $\hat{w}_1 = w_1 z(-\ln p_*) - N + 1$  and  $\hat{w}_2 = w_2 z(-\ln p_*) - N + 1$ . The coefficient in (11) is equal to

$$\begin{aligned} & \frac{(-\ln p_*)}{z} \times \left( p_*^{w_1 z} \left( w_1 z - \frac{N-1}{(-\ln p_*)} \right) \right. \\ & \quad \left. - p_*^{w_2 z} \left( w_2 z - \frac{N-1}{(-\ln p_*)} \right) \right) \\ &= \frac{(-\ln p_*)}{z} \times (\hat{w}_1 \exp(-\hat{w}_1) - \hat{w}_2 \exp(-\hat{w}_2)). \end{aligned}$$

Therefore,

$$\partial F/\partial M_1 > 0 \iff \hat{w}_1 \exp(-\hat{w}_1) > \hat{w}_2 \exp(-\hat{w}_2). \quad (12)$$

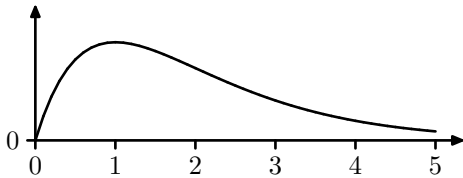


Figure 1: The graph of the function  $x \exp(-x)$

The right-hand side of (12) has no explicit solution, but it can be solved numerically using the Lambert function ([19], [11]). The function  $x \exp(-x)$  increases in the interval  $(0, 1]$  and decreases in the interval  $[1, \infty)$  (see Figure 1). Since  $w_1 < w_2$ , we have  $\hat{w}_1 < \hat{w}_2$ . Therefore we have

$$\hat{w}_1 \geq 1 \implies \frac{\partial F}{\partial M_1} > 0. \quad (13)$$

By the definition of  $\hat{w}_1$ , a sufficient condition for  $\hat{w}_1 \geq 1$  is

$$z \geq \frac{N}{(-\ln p_*)w_1}.$$

Since  $F$  is decreasing in  $z$ , and since  $z$  is a solution of (8),  $z \geq N/(-\ln p_*)w_1$  if and only if

$$F\left(w_1, w_2, M_1, M_2, \frac{N}{(-\ln p_*)w_1}\right) \geq \frac{dN}{T\mu}.$$

Observe that

$$F\left(w_1, w_2, M_1, M_2, \frac{N}{(-\ln p_*)w_1}\right) \approx \frac{N-1}{N}(-\ln p_*)w_1.$$

Indeed, all terms in the definition of  $F$  that contain an exponent are small when  $N$  is large:  $p_*^{w_1 N/w_1(-\ln p_*)} = \exp(-N)$  and  $p_*^{w_2 N/w_1(-\ln p_*)} = \exp(-w_2 N/w_1) < \exp(-N)$ . Thus, when  $N$  is large, if  $w_1 \geq dN^2/(N-1)T\mu(-\ln p_*)$  then  $\partial F/\partial M_1 > 0$ , which is equivalent to  $\partial z/\partial M_1 > 0$ .  $\square$

Now we study the effect of the registry on the efficiency of the internet. We show that as more type-1 users join the registry, the total number of spam messages sent by spammers decreases. Thus, to increase efficiency one should encourage users to join the registry. Type-2 users have a similar effect, provided the cost of sending spam mail is not too high. This effect is summarized by the following two theorems.

**THEOREM 12.** *Assume that  $N \geq 2$  and  $w_1 < w_2$ . The function  $M_1 \mapsto (M_1 + M_2) \times z(w_1, w_2, M_1, M_2, d)$  is monotonic increasing: as  $M_1$ , the number of type-1 users who do not join the registry, decreases, the total number of spam messages sent decreases as well.*

PROOF. Denote by  $v = (M_1 + M_2) \times z$  the total number of spam messages sent by all the spammers. Define

$$\begin{aligned} G(M_1, M_2, v) &= F(w_1, w_2, M_1, M_2, \frac{v}{M_1 + M_2}) \\ &= (-\ln p_*) \left( \frac{M_1}{M} w_1 p_*^{w_1 v/M} + \frac{M_2}{M} w_2 p_*^{w_2 v/M} \right) \\ &\quad + \frac{N-1}{v} \left( M_1 \left( 1 - p_*^{w_1 v/M} \right) + M_2 \left( 1 - p_*^{w_2 v/M} \right) \right). \end{aligned}$$

Since the function  $v \mapsto p_*^{c v}$  is monotonic decreasing for every  $c > 0$ , and since the function  $v \mapsto (1 - p_*^{c v})/v$  is monotonic decreasing for every  $c > 0$  we have  $\frac{\partial G}{\partial v} < 0$ .

Set  $H(M_1) = G(M_1, M_2, Mz(w_1, w_2, M_1, M_2, d))$ . By (8), we have  $H(M_1) = \frac{dN}{T\mu}$  for every  $M_1$ , hence  $H'(M_1) = 0$ . By the chain rule,

$$0 = H'(M_1) = \frac{\partial G}{\partial M_1} + \frac{\partial G}{\partial v} \frac{\partial v}{\partial M_1}.$$

Since  $v = Mz$ , we need to prove that  $\partial v/\partial M_1 > 0$ . Since  $\partial G/\partial v < 0$ , it is sufficient to show that  $\partial G/\partial M_1 > 0$ .

The derivative of  $G$  w.r.t.  $M_1$  is

$$\frac{\partial G}{\partial M_1} = (-\ln p_*) \frac{M_2}{M^2} \left( w_1 p_*^{w_1 v/M} - w_2 p_*^{w_2 v/M} \right) \quad (14)$$

$$+ (-\ln p_*)^2 \frac{M_1 w_1^2 v}{M^3} p_*^{w_1 v/M} \quad (15)$$

$$+ (-\ln p_*)^2 \frac{M_2 w_2^2 v}{M^3} p_*^{w_2 v/M} \quad (16)$$

$$+ \frac{N-1}{v} \left( 1 - p_*^{w_1 v/M} \right) \quad (17)$$

$$- (N-1) \frac{w_1 M_1}{M^2} (-\ln p_*) p_*^{w_1 v/M} \quad (18)$$

$$- (N-1) \frac{w_2 M_2}{M^2} (-\ln p_*) p_*^{w_2 v/M}. \quad (19)$$

Since  $w_1 < w_2$  we have  $-p_*^{cw_1} > -p_*^{cw_2}$  for every  $c > 0$ , and therefore the term (17) can be split as follows.

$$\frac{N-1}{v} \left( 1 - p_*^{w_1 v/M} \right) = \quad (20)$$

$$= \frac{M_1}{M} \frac{N-1}{v} \left( 1 - p_*^{w_1 v/M} \right) \quad (21)$$

$$+ \frac{M_2}{M} \frac{N-1}{v} \left( 1 - p_*^{w_1 v/M} \right) \quad (22)$$

$$\geq \frac{M_1}{M} \frac{N-1}{v} \left( 1 - p_*^{w_1 v/M} \right) \quad (23)$$

$$+ \frac{M_2}{M} \frac{N-1}{v} \left( 1 - p_*^{w_2 v/M} \right). \quad (24)$$

By Lemma 14,  $1 - p_*^x > x(-\ln p_*)p_*^x$  for every  $x > 0$ , so that the difference between the term in (23) and the term in (18) is positive. We will now show that the sum of the terms (14), (16), (18), (24) is non-negative. This will conclude the proof.

Indeed, setting  $w = w_2/M$ , the sum is (we omitted the quantity  $M_2/M$  which multiplies all terms):

$$\begin{aligned} A(w) &= (-\ln p_*) w p_*^{wv} + (-\ln p_*)^2 w^2 v p_*^{wv} \\ &\quad - (N-1) w (-\ln p_*) p_*^{wv} + \frac{N-1}{v} (1 - p_*^{wv}). \end{aligned}$$

A simple calculation shows that the derivative is non-negative as soon as  $N \geq 2$ . Since  $A(0) = 0$ , we deduce that  $A(w) \geq 0$  for every  $w$ , as desired.  $\square$

**THEOREM 13.** *Provided  $N$  is large, if  $w_2 > \frac{dN^2}{(N-1)T\mu(-\ln p_*)}$  the function  $M_2 \mapsto (M_1 + M_2) \times z(w_1, w_2, M_1, M_2, d)$  is monotonic increasing: as  $M_2$ , the number of type-2 users who do not join the registry, decreases, the total number of spam messages sent decreases as well.*

**PROOF.** Assume that  $N$  is large, and that  $w_2 > \frac{dN^2}{(N-1)T\mu(-\ln p_*)}$ . By Theorem 11 we have  $\frac{\partial z}{\partial M_2} > 0$ . In other words, the function  $M_2 \mapsto z(M_1, M_2)$  is increasing. Therefore the function  $M_2 \mapsto (M_1 + M_2) \times z$ , as a product of two increasing functions, is increasing as well.  $\square$

## 5. FINAL COMMENTS

The model we presented is simplistic, and further research is needed to improve our understanding of the optimal behavior of spammers in real life.

The few assumptions we made that do not change our conclusions are that there are only two types of users, and that a user who decides to purchase a spam product does

so only once, and spends a fixed amount on this purchase. Allowing for a more diversified environment will only make the calculations more complex. As we mentioned in the text, the assumption that the spammers send the same number of messages to all users is also not crucial for our results.

An important point we ignored is the behavior of the anti-spam companies. For example, anti-spam companies adapt their filters to the techniques used by spammers. A spammer who increases the number of messages he or she sends has a higher chance of having the filters learn how to identify his or her messages. This effect may reduce the number of spam messages sent by spammers. Indeed, anti-spam companies spend their efforts on adapting their filters to common techniques rather than to unusual ones. As a consequence, the spammers and the anti-spam companies play a game, in which spammers have to spend time on developing new techniques, and determining how often to use each technique, and anti-spam companies have to decide which technique should be dealt with first. Studying this game will improve our understanding of the spam market.

Another possible extension of the model is to allow new spammers to enter the market, and existing spammers to leave it.

Filters and payments are not the only way to fight spam mail. Authentication and reputation services are another important set of technologies. It would be interesting to develop a model that includes these technologies and study their effects on spam volume.

Finally, it will be interesting to combine our model with the model presented in [3], and analyze a game between many spammers and many users who can tune their filters, and the effect of improving the quality of the filters on the behavior of both spammers and users.

## 6. REFERENCES

- [1] M. Abadi, A. Birrell, M. Burrows, F. Dabek and T. Wobber (2003) Bankage Postage for Network Services. *Advances in Computing Science ASIAN 2003, Lecture Notes in Computer Science #2896*, 72-90, Springer-Verlag.
- [2] S. Ahmed and F. Mithun (2004) Word Stemming to Enhance Spam Filtering. *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS 2004)*, Mountain View, CA, USA.
- [3] I. Androutsopoulos, E.F. Magirou and D.K. Vassilakis (2005) A Game Theoretic Model of Spam E-Mailing. *Forthcoming, CEAS 2005*.
- [4] I. Androutsopoulos, G. Paliouras, V. Karkaletsis, G. Sakkis, C.D. Spyropoulos and P. Stamatopoulos (2000) Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach. In H. Zaragoza, P. Gallinari, and M. Rajman (Eds.), *Proceedings of the Workshop on Machine Learning and Textual Information Access, 4th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2000)*, Lyon, France, pp. 1-13.
- [5] X. Carreras and L. Marquez (2001) Boosting Trees for Anti-Spam Email Filtering. *Proceedings of RANLP-01, 4th International Conference on Recent Advances in Natural Language Processing*.
- [6] Z. Chuan, L. Xianliang and X. Qian (2004) A Novel Anti-spam Email Approach Based on LVQ. *Proceedings of the 5th International Conference, PDCAT 2004, Singapore*. Eds. Liew K.-M., Shen H., See S.
- [7] J. Clark, I. Koprinska and J. Poon (2003) A Neural Network Based Approach to Automated E-Mail Classification. *Proceedings of the IEEE/WIC International Conference on Web Intelligence*, p.702.
- [8] L.F. Cranor and B.A. LaMacchia (1998) Spam! *Communications of the ACM*, **41**, 74-83. <http://doi.acm.org/10.1145/280324.280336>
- [9] N. Dalvi, P. Domingo, Mausam, S. Sanghai and D. Verma (2004) Adversarial Classification. *Proceedings of the Tenth ACM SIGKDD International conference on Knowledge Discovery and Data Mining (KDD'04)*, 99-108.
- [10] C. Dwork and Naor M. (1993) Pricing via Processing or Combatting Junk Mail. *Advances in Cryptology - CRYPTO 1992, Lecture Notes in Computer Sciences #740*, 139-147.
- [11] L. Euler (1783) De serie Lambertina Plurimisque eius insignibus proprietatibus. *Acta Acad. Scient. Petropol.*, **2**, 29-51, 1783. Reprinted in L. Euler, *Opera Omnia, Ser. 1, Vol. 6: Commentationes Algebraicae*. Leipzig, Germany: Teubner, pp. 350-369, 1921.
- [12] E. Gabber, M. Jakobsson, Y. Matias, and A. Mayer (1998) Curbing junk E-mail via secure classification. In *Proceedings of Financial Cryptography '98*, Anguilla, BWI.
- [13] F.D. Garcia, J.H. Hoepman and J. van Nieuwenhuizen (2004) Spam Filter Analysis. *Proceedings of 19th IFIP International Information Security Conference, WCC2004-SEC*
- [14] A. Gray and M. Haahr (2004) Personalised, Collaborative Spam Filtering. *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS 2004)*, Mountain View, CA, USA.
- [15] R.J. Hall (1998) How to Avoid Unwanted Email. *Communications of the ACM*.
- [16] D.W.K. Khong (2004) An Economic Analysis of Spam Law. *Erasmus Law and Economics Review*, **1**, 23-45.
- [17] R.E. Kraut, S. Sunder, J. Morris, R. Telang, D. Filer and M. Cronin (2002) Will Postage for Email Help? *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work (CSCW'02)*, 206-215.
- [18] R.E. Kraut, S. Sunder, R. Telang, and J. Morris (2005) Pricing Electronic Mail To Solve the Problem of Spam. *Human-Computer Interaction Vol. 20*, 195-223.
- [19] J.H. Lambert (1758) *Observations variae in Mathesin Puram. Acta Helvetica, physico-mathematico-anatomico-botanico-medica*, **3**, 128-168.
- [20] B. Laurie and Clayton R. (2004) Proof-of-Work Proves not to Work. *Workshop on Economics and Information Security*.
- [21] B. Leiba and N. Borenstein (2004) A Multifaceted Approach to Spam Reduction. *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS 2004)*, Mountain View, CA, USA.
- [22] K. Li, C. Pu and M. Ahamad (2004) Resisting SPAM Delivery by TCP Damping. *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS 2004)*, Mountain View, CA, USA.
- [23] T. Loder, M. Van Alstyne and R. Wash (2004) Information Asymmetry and Thwarting Spam. *ACM Electronic Commerce*.
- [24] E. Michelakis, I. Androutsopoulos, G. Paliouras, G. Sakkis and P. Stamatopoulos (2004) Filtron: A Learning-Based Anti-Spam Filter. *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS 2004)*, Mountain View, CA, USA.
- [25] P. Pantel and D. Lin (1998) SpamCop- A Spam Classification & Organization Program. *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*.
- [26] I. Rigoutsos and T. Huynh (2004) Chung-Kwei: a Pattern-discovery-based System for the Automatic Identification of Unsolicited E-mail Messages (SPAM). *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS 2004)*, Mountain View, CA, USA.
- [27] M. Sahami, S. Dumais, D. Heckerman and E. Horvitz (1998) A Bayesian Approach to Filtering Junk E-Mail. *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*.
- [28] R. Singel, Under Attack, Spam Fighters Folds. *Wired News*, May-16-2006.
- [29] T. Tompkins and D. Handley (2003) Giving E-mail Back to the Users: Using Digital Signatures to Solve the Spam Problem. *First Monday*, **8**.
- [30] T. Van Zandt (2004) Information Overload in a

Network of Targeted Communication. RAND Journal of Economics, **35**, 542-560.

- [31] W. Yeraunis (2003) Sparse Binary Polynomial Hash Message Filtering and The CRM114 Discriminator. Proceedings of 2003 MIT Spam Conference.

## 7. ACKNOWLEDGMENTS

We thank Ion Androutsopoulos, Nathaniel Borenstein, Jeff Kaphart, seminar participant's at IBM research labs, and three anonymous referees, for their insightful comments.

## Appendix

Here we prove some technical results we need regarding the function  $(1 - p^x)/x$ .

LEMMA 14. *For every  $p \in (0, 1)$  and every  $x > 0$  one has*

$$x(-\ln p)p^x < 1 - p^x < (-\ln p)x. \quad (25)$$

PROOF. Since all terms in Eq. (25) vanish at  $x = 0$ , it is sufficient to compare their derivatives. That is, we need to prove that

$$(-\ln p)p^x - x(-\ln p)^2 p^x < (-\ln p)p^x < (-\ln p).$$

However, this inequality holds since  $p \in (0, 1)$  and  $x > 0$ .  $\square$

By L'hospital's rule, one obtains the following:

$$\lim_{x \rightarrow 0} \frac{1 - p^x}{x} = -\ln p. \quad (26)$$

LEMMA 15. *For every  $p \in (0, 1)$  and every  $q > 0$ , the function  $f(z) = \frac{1 - p^{qz}}{z}$  is monotonic decreasing over  $(0, \infty)$ .*

PROOF. The derivative of  $f$  is

$$\begin{aligned} f'(z) &= -\frac{1}{z^2}(1 - p^{qz}) + \frac{1}{z}(-\ln p)qp^{qz} \\ &= \frac{1}{z^2}(qz(-\ln p)p^{qz} - (1 - p^{qz})). \end{aligned}$$

By Lemma 14, we have  $1 - p^{qz} > qz(-\ln p)p^{qz}$ , and the result follows.  $\square$