Modelling Incentives for Email Blocking Strategies

Andrei Serjantov and Richard Clayton

The Free Haven Project, UK
schnur@gmail.com
University of Cambridge, Computer Laboratory,
JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom
richard.clayton@cl.cam.ac.uk

Abstract. We consider the current arrangements whereby many users send their email via ISP "smarthosts" and model this system as an economic network. Some of these users will have insecurely configured machines, which may be hijacked for the sending of spam. This can lead to other sites blocking all email from the smarthost, which will then affect all of the ISP's users. We derive formulae that express when blocking is an appropriate action and show that the number of legitimate users at the ISP must be taken into account as well as the number of spam sources. We relate this formal model to real world behaviour and attitudes and generate some strategies that maximise utility for different sizes of ISP. We present some quantitative email data from a large ISP that indicates the likely range in which numerical solutions will lie. We also demonstrate that some of the present approaches to email blocking are irrational and counterproductive.

1 Introduction

Many ways of filtering out unwanted bulk email (spam) have been proposed [11]. Content-based methods are often effective, but computationally expensive when done at the Internet Service Provider (ISP) where volumes are high [1]. A cheap and reasonably effective method of filtering is simply to refuse email from particular sources on the basis that they send nothing but spam. Many people delegate the blocking of spam to their ISP, sometimes without actually knowing that they are doing so [4]. Some ISPs compile blacklists to suit their own purposes, but many delegate their decisions by adopting the recommendations of public blacklists, such as those from Spamhaus [13]. There are now well over 600 "blacklists"[9] that can be consulted when making blocking decisions, though the reasons given for inclusion on these lists varies widely – as do the policies for any subsequent removal. In late 2004 it was alleged that Verizon Inc had blacklisted all mail servers outside the USA which led to court action [7]. It is the use of blacklisting as a spam filtering (blocking) tactic and, in particular, the criteria being applied by ISPs that are the subject of this paper.

We argue that each ISP should add hosts to a blacklist only after giving due consideration to the benefit of doing so. We build a very simple model which

Workshop on the Economics of Information Security, WEIS05, 2–3 June 2005

allows us to show how this should be done. We then present some evidence (based on interviews with Spamhaus and with Demon Internet, a large ISP in the UK) that our model is able to predict and explain some real world behaviour. Our model highlights some particular features of the email blocking problem, which, although rather obvious, have not previously been clearly identified. We believe that our model can assist ISPs in *rationally* deciding whether to block a particular host or not.

2 The Model

Following the approach presented in Jackson [6], we model the email network as a graph with a set of nodes N and a set of edges $E \subseteq N \times N$. Nodes in this graph represent hosts that send out email – these could be ISP smarthosts (email servers used by ISP customers as a relay for their outgoing email) or corporate email servers (that send their email directly, bypassing their ISP's systems).

Each node sends email on behalf of a collection of users, who are vulnerable to attacks from viruses, worms, trojan horses, etc. and hence may be unwittingly sending out spam. Of course there will also be some nodes that are owned by spammers and hence their sending of spam is entirely intentional – but only one third of all spam is now sent directly from spammers, and this proportion is rapidly decreasing [3].

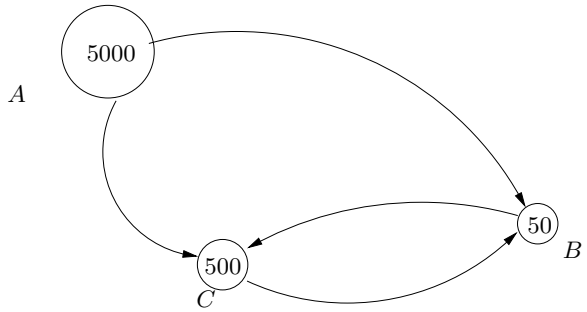

Fig. 1. An Email Network

An example of an email network is shown in Figure 1. In the diagram the arrows express the flows of emails which are permitted within the network. Hence in this example, the large ISP A is not happy to receive email from either B or C, whereas they are both prepared to swap email with each other and accept it from A.

The utility of each of the hosts of the network is a function of its connectivity (who email can be sent to and who it can be received from), but this utility is reduced by the amount of spam being received from remote sites. Clearly any

particular user is happiest if she is able to send email to and receive email from as many people as possible, while receiving the least possible amount of spam.

We use C_A to denote the set of users (clients) belonging to node A and a function $V_A : C_A \to \mathbb{N}$ to represent the cost of receiving the spam sent by A's users (due to intentional sending, or because some of them have been compromised by the various attacks) to one random email address. Essentially, V_A measures how vulnerable the users C_A are. We can then express the utility of a node as follows:

$$U(A) = \left(\sum_{B \text{ s.t. } (A,B) \in E} U(C_B) \right) - \left(|C_A| \times \sum_{B \text{ s.t. } (B,A) \in E} V_B(C_B) \right)$$

where the first, positive, term involves $U(C_B)$, the utility of being able to send email to the clients of node B and the second, negative, term involves $V_B(C_B)$, the vulnerability of the clients C_B (the likelihood they will send spam) which we then multiply up by all the clients C_A of A that they might send it to.

However, it is not only the vulnerability of clients which might cause A to block the flow of email from B. If the incentives of node A (the ISP, the smarthost, the corporation running the email server) are different to those of $a \in C_A$, i.e. the users of the node, then blocking could take place contrary to the wishes of the users. One such example is corporations blocking email flow to and from free email services such as Hotmail or Gmail due to "too much personal use of company resources".

One might expect to see a further positive term in the formula for $U(A)$ to express the utility of being able to receive email from remote sites, but we do not include it. The reason is that there is significant anecdotal evidence suggesting that whether users are able to receive email or not is irrelevant to the overall utility function. Users rarely learn about the email they do not receive, and even if they do, are seldom competent enough to assign the blame to their ISP blocking it rather than the sender having made an error.

Of course, a great deal of email will not actually travel to remote ISPs but will be delivered locally. ISPs tend to serve small geographical areas and this, along with their market positioning, will mean that they have similar types of clientele who are likely to exchange email amongst themselves. Also, the very large global entities, such as AOL or MSN, send a lot of email internally because of their sheer size. We can therefore expect the self-sending terms of the utility function:

$$U(A)_{\text{self}} = U(C_A) - |C_A| \times V_A(C_A)$$

to be especially significant. These terms may be even larger because each ISP will be in an excellent position to act against their own users sending spam to each other and hence the negative term will tend to zero. So we can expect one of the largest terms in the utility function to be:

$$U(A)_{\text{self}} = U(C_A)$$

One might reasonably argue that summing utility values is merely a lower bound to the total utility and that one might expect to see a power function (following "Metcalfe's Law", this would be squaring[1]) although recent work by Odlyzko argues for an $n \log(n)$ valuation [10].

Of course, each of the nodes has complete control over whether it will receive email from other nodes, i.e. over the edges coming into it. Hence, if nodes are behaving rationally they will accept email from all nodes where this increases overall utility, viz: this means that node A should have an edge to node B when $U(C_B) - |C_A| \times V_B(C_B) \geq 0$. Quite clearly, where the "vulnerability" is complete (viz: the node is owned by spammers and sends nothing but spam) then the second term will be large. At the same time it is most unlikely that there will be any wish to send email back to the spammers and so the overall utility is bound to be negative and blocking is the rationally correct strategy.

3 Implications of the Blocking Model

The model shows one way to rationally decide whether to block email from a particular node or not. This has several implications, which we relate to situations which have arisen in practice.

First of all, ISPs are adversely affected by their customers' security vulnerabilities. Some ISPs already recognise this and take a proactive approach, scanning their customers' machines for problems and notifying them that patches need to be applied. This is expensive, but it reduces the possibility of customers' machines being compromised (represented in our model by V_A), spam being seen to come from the ISP's smarthost, and the smarthost being blocked.

Clearly, the model also implies that "secure" customers are more valuable than insecure ones. For instance, customers running operating systems and email clients which are not vulnerable to standard attacks are much less likely to be compromised and thereby cause spam to be sent through the ISP smarthost. In practice it is the presence of insecure or actively spamming customers that is usually acted upon by other ISPs. According to reports [8], a high percentage of traffic coming from one large UK ISP constitutes "advanced fee fraud" scams, which causes all of its email to be blocked by many other ISPs. Another account was of a large US ISP that never terminated accounts used by spammers. Hence, many of its smarthosts became blocked and customers were unable to send email (interestingly, the company policy only changed when the corporate email server was blocked, thus preventing executives from sending email!).

Spam is sent in two distinct ways, scattershotted to a selection of probably valid email addresses, or as an indiscriminate "dictionary attack" that attempts to find deliverable addresses by random guessing. In the latter case, the size of receiving ISP is irrelevant, but in the former case, the larger the ISP, the more likely it is to receive spam and the more likely it is that there will be users who notice. This means that there will be a tendency for larger ISPs to be far more

[1] "the power of the network, how much it can do, is the square of the number of connected machines", Robert Metcalfe, 1973

aware of spam sources and hence they will tend to block more. This seems to be supported by evidence in practice (Demon seems to be regularly blocked by one careless large US ISP [5]). In particular, at a very large ISP, since the self-sending term is so large, almost all the other terms will be "lost in the noise" and it may be that many users are entirely, or almost entirely, unaware of extensive blocking. Conversely, a small ISP will see little benefit from blocking and its utility function will be dominated by the ability to send to others.

We have argued that utility is unaffected by whether or not users receive email, this could mean "tit for tat" blocking is a useful strategy. If A has decided to block all email arriving from B then B might consider blocking incoming email from A in retaliation on the basis that the users C_B will not notice. However, whether A takes any notice will depend on the size of the two terms in the model, which will significantly affected by the relative sizes of the two entities. If C_A is large compared with C_B then A may just not care that they have been blocked, however if B is the larger entity then it may be that "tit for tat" will be effective in getting the block removed.

3.1 The ISP Experience

Demon Internet told us that their reputation was of considerable importance. viz: that other ISPs would not block them immediately but would give them a chance to fix things because they had managed such fixes in the past. This corresponds to the other ISPs being unable to measure $V_B(C_B)$ but relying on estimates of its likely size. An obvious corollary of this is that it is important for an ISP to be proactive in publicising its effectiveness at dealing promptly with spam coming from its systems. This will make it less likely to be blocked because of the over-optimistic view that others will then take of the vulnerability term.

Demon's experience is that email blocking is often put in place for long periods. Some sites are quick to block and take a significant time to unblock, even when the original problem has been fixed. This is irrational because it uses an overly pessimistic view of the vulnerability term, but our simple model does not include any costs for making decisions about blocking or unblocking and implementing them; and here we are probably seeing a lack of resources for making frequent changes.

4 Quantitative Effects of Email Being Blocked

In this section we consider the quantitative effect on an ISP of having outgoing email blocked. In terms of the model, we are examining the nature of $U(C_B)$ the utility of being able to send email to some other node B.

4.1 Experimental Results

We were able to examine traffic logging data for outgoing email from the Demon Internet smarthost for 15 Feb 2005 to 14 Mar 2005, a one month (28 day) period

that did not include any UK holidays. Using the methods outlined in [2] we excluded 498 customers who had been detected to be infected with email viruses or worms, along with those who were insecure and were being exploited to send spam. We also excluded a further 56 who were looping emails through the system (using a cut-off of > 2000 looping items on any one day). The remaining data for 82 062 customers related to 25 245 000 emails. However, 9 857 191 of these emails had a null sender – mainly delivery failure reports for incoming spam – and these were ignored. This left 15 387 809 emails addressed to 21 687 885 destinations.

The destinations were analysed by the destination mail server for each domain (viz: the DNS was consulted by looking up the MX record, or failing that the A record). Where there were multiple servers, we made a consistent choice. We found that there were 378 271 destinations, but further inspection showed that 240 850 were only used once during the three week period – the overwhelming majority were associated with spam rejection and the rest appeared to be mainly typographical errors. The top 20 destinations are given in Table 1.

Destination	Role	Emails	Customers
messagelabs.com	Spam filtering	1 361 916	35 641
hotmail.com	Global webmail	1 320 900	43 350
aol.com	Global ISP	820 645	37 674
btinternet.com	UK ISP	809 367	39 048
yahoo.co.uk	UK portal	367 327	24 302
demon.net	UK ISP (self)	363 112	15 212
ntl.com	UK ISP	337 441	25 174
yahoo.com	Global portal	298 491	18 139
uk.tiscali.com	UK ISP	235 858	22 022
virgin.net	UK ISP	189 389	18 358
schlund+partner	German web hosting	166 077	14 540
nhs.uk	UK health service	160 793	9 816
blueyonder.co.uk	UK ISP	149 521	14 677
pipex.net	UK ISP	97 495	9 576
spicerhaart.co.uk	UK estate agent	85 425	144
clara.net	UK ISP	82 309	8 106
mailcontrol.com	Spam filtering	80 941	6 978
global.net.uk	UK ISP	77 957	10 586
plus.net	UK ISP	77 080	9 289
postini.com	Spam filtering	74 777	6 092

Table 1. Top 20 Destinations of Email Sent From Demon Internet

It is noteworthy that the highest volume destination was `messagelabs.com` who operate mail servers on behalf of other organisations, remove any spam and then forward the email to its true destination. It seems reasonable to assume that a consistent blocking policy is being adopted for all of these disparate organisations. Also, $U(A)_{\text{self}}$ (i.e. email sent from one Demon Internet customer

to another) was, as predicted, a large value – albeit it was only at #6 in the table. One might expect AOL, Hotmail (which also handles MSN email) or Yahoo! to have this component at #1 in their own list of destinations.

Figure 2 graphically shows how this distribution continues. We have sorted the sites by the number of emails sent to them and then plotted this, on a log scale, against the cumulative number of emails. To avoid edge effects we have discarded the data for sites where fewer than 10 customers sent email. As can be seen the graph indicates that there is a power law operating. That is to say, the amount of email sent to the top few sites is about the same as that sent to the next 10 sites, which is in turn about the same as that sent to the next 100 and that is again about the same as is sent to the next 1000 and so on.

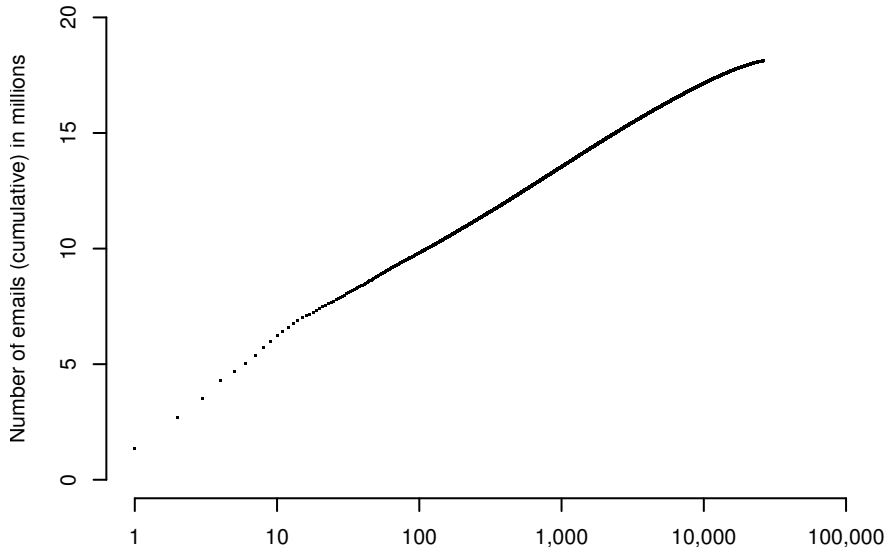

Fig. 2. Email Destinations

Another way of viewing the data is in Figure 3, where we consider how many sites have large numbers of customers sending email. Thirteen sites have over 10 000 customers sending to them, 213 sites more than 1 000 customers and 2 601 sites would affect more than 100 Demon Internet customers if they were to block incoming email.

What these figures mean in practice is that it is not possible for Demon to concentrate its efforts in ensuring that only a small number of sites do not block their email. They need to ensure that many thousands of sites allow their customers' emails through – and should any of these sites block email then hundreds, if not thousands, of customers will be inconvenienced.

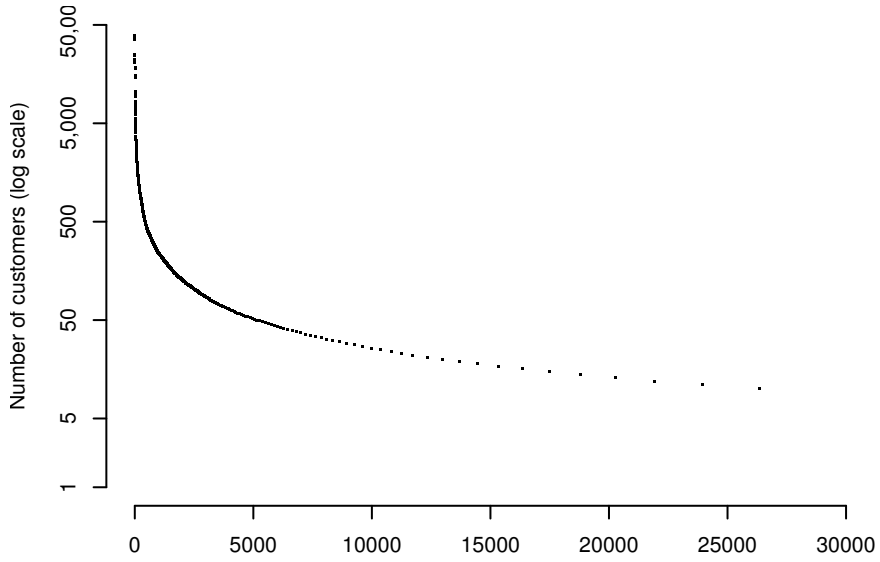

Fig. 3. Destination Popularity

5 Towards a Rational Blocking Strategy

Unfortunately, the utility function given in Section 2 does not readily suggest a decision procedure as to whether an ISP should block a particular smarthost or not. Also, as we suggested, the interests of an ISP are not always perfectly aligned with the interests of their customers. Notwithstanding, in this section we will depart from our idealistic stance and look at the problem from the ISP's point of view – which smarthosts should they block?

To investigate the way in which a blocking policy decision could be made we examined logging data for Demon Internet's incoming email between 26 April and 9 May 2005 (14 days). Approximately 55.6 million emails were received, of which 66.5% were identified as spam by Demon's filtering system. We then looked up the IP address from which each email came and established the source AS (Autonomous System), viz: which of 13 378 ISPs was responsible for the sending system(s). Data is not available to establish which IP addresses within ASs correspond to smarthosts; but ISP blocking policies are likely to be implemented at the subnet or AS level, so the distinction between AS and smarthost is unlikely to be relevant in practice.

The data is very complex. In general, the more spam that arrives, the more non-spam accompanies it. This does of course reflect the varying sizes of ISPs, but the accuracy of the spam/non-spam discrimination is clearly relevant.

Any email blocking policy that aims to reduce the amount of incoming spam will clearly look first at ASs which send few non-spam messages. We therefore

examined the frequency distributions for spam email accompanied by up to 9 non-spam emails and plot these ten distributions in Figure 4. As can be seen, there is much commonality for all these small numbers of non-spam emails.

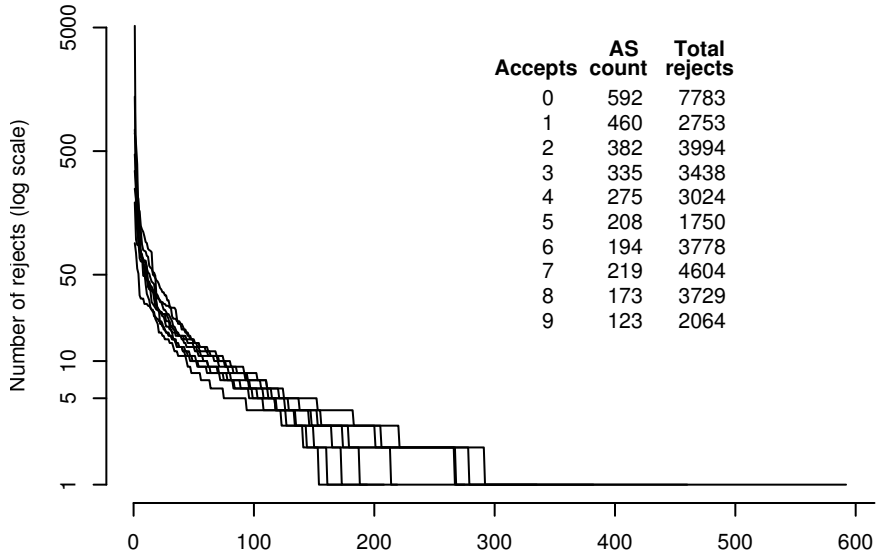

Fig. 4. Frequency Distribution of Spam per AS for Small Values of Non-Spam Email

Further examination of the rest of the data shows a plausible approach to deciding whether or not to block an AS. Considering solely the 6 547 ASs which sent 28 or fewer non-spam emails during the two weeks, there are 14 which also sent more than 560 spam emails each. Blocking these sources, which would be an acceptable amount of work to implement, would have resulted in 21 945 spam emails being blocked (along with 165 non-spam).

We do not claim that this blocking would necessarily be worthwhile, nor that the particular values we have chosen are in any way optimal – there are complex trade-offs here that we do not wish to dwell upon (and of course any definite advice would need to be based on studying much more than two weeks data at one ISP). Nevertheless, we believe that this type of strategy is worthy of further study; and in particular a 24-hour sample of logging data from a few weeks earlier contained a single AS which, had it been blocked, would have automatically rejected 9 948 emails every single one of which was identified as spam. This same AS looks harmless in our two-week sample (40 spam emails, 34 non-spam) which suggests that blocking policies have to be extremely dynamic, and react quickly to changes in email volumes.

In passing, we can note that it is easy to use the data to show the folly of much more simplistic blocking strategies. For example, if Demon simply blocked the 100 ASs that sent them the most spam then this would discard about three-quarters of all their incoming spam; however it would also result in the blocking of about half a million non-spam messages per day, which is unlikely to be an acceptable result for them.

It is also interesting to compute the value of a decision to block an AS. Demon estimate that the cost of their filtering solution averages out to about 0.04p per rejected email. Hence any policy that succeeded in rejecting 10 000 emails a day would account for about £1500/year. Of course this number is almost meaningless – the marginal cost of identifying this spam is practically zero; and the filtering system is monolithic and step changes of investment come in lumps of tens of thousands of pounds. Nevertheless, with sufficient blocking then one might be able to stave off future enhancements to the system and save money that way.

All of this, however, is really somewhat speculative and much more work on real world data is needed. Our current opinion can be summarised by saying that although we are not currently prepared to recommend any particular blocking strategies, we continue to believe that it is possible to develop such strategies in a rational manner and put forward economic justifications for adopting them.

6 Conclusions

In this paper we have outlined an initial approach to understanding the economics of ISPs using email blocking strategies to combat incoming spam email. We pursued two directions – designing a simple model for this problem and trying to find some evidence supporting it by conducting interviews at a UK ISP and Spamhaus.

From this we gained an insight into this very practical problem and this has allowed us to propose a rational way for ISP to implement email blocking policies. We have also documented a number of motivations for blocking which are outside the model and may be overriding in their decision making. Although these have been observed in practice, they are not general enough to apply to every ISP and hence are not incorporated in the model.

Our access to some quantitative data means that we have been able to assess the scale of the problem facing ISPs. The very long "tail" to the distribution of destinations means that it is not practical to run on the basis of a handful of "special relationships", but ISP policies must ensure that many thousands of destinations are prepared to accept their email.

Although we have presented some initial thoughts as to the basis on which such decisions could be made, we are not yet able to make clear recommendations as to whether it is rational for ISP A to block ISP B's smarthost. However, by reporting some of the most interesting observations arising from our work so far we hope to stimulate further research in this area.

Acknowledgments

The authors wish to thank Steve Linford at Spamhaus and James Hoddinott and Alex Kiernan at Demon Internet for their assistance in understanding the real-world dynamics of email blocking. We also acknowledge the financial assistance provided by the Cambridge MIT Institute (CMI) to Richard Clayton through the project: "The design and implementation of third-generation peer-to-peer systems".

Disclaimer: Nothing in this paper should be read as being a statement of Demon Internet's current policy on email blocking or as an indication of any future policy decisions.

References

1. Aalto J.(ed): Procmail tips. http://pm-doc.sourceforge.net/pm-tips.html
2. Clayton R.: Stopping Spam by Extrusion Detection. First Conference on Email and Anti-Spam (CEAS 2004), Mountain View CA, USA, Jul 2004.
3. Gaudin S.: Viruses Blamed for Expected 80% Spam Saturation by Q3. Datamation, 24 Feb 2004. http://itmanagement.earthweb.com/secu/article.php/3317271
4. Global Internet Liberty Campaign (GILC) and the Internet Free Expression Alliance (IFEA): Coalition statement against "stealth blocking", 17 May 2001. http://www.gilc.org/speech/stealth_blocking.html
5. Hoddinott J., Demon Internet, personal communication.
6. Jackson M.: Allocation Rules for Network Games Working Paper, California Institute of Technology, Mar 2003.
7. Leyden J.: Verizon faces lawsuit over email blocking. The Register, 21 Jan 2005. http://www.theregister.co.uk/2005/01/21/verizon_class_action
8. Linford S., Spamhaus, personal communication.
9. Makey J.: Blacklists Compared. 3 Apr 2004. http://www.sdsc.edu/~jeff/spam/cbc.html
10. Odlyzko A. and Tilly B., A refutation of Metcalfe's Law and a better estimate for the value of networks and network interconnections. 2 Mar 2005. http://www.dtc.umn.edu/~odlyzko/doc/metcalfe.pdf
11. Postini Inc: Expert's Guide to Choosing an Anti-Spam Solution. Postini Inc White Paper, 31 Mar 2004. http://www.postini.com/whitepapers/WP05-01-0403_Postini_Experts_Guide.pdf
12. Sahami M., Dumais S., Heckerman D. and Horvitz E.: A Bayesian Approach to Filtering Junk E-mail. AAAI'98 Workshop on Learning for Text Categorization, 27 Jul 1998, Madison, Wisconsin.
13. http://www.spamhaus.org/