

An Overview of Spam Handling Techniques

Prashanth Srikanthan
Computer Science Department
George Mason University
Fairfax, Virginia 22030

Abstract

Flooding the Internet with many copies of the same email message is known as spamming. While spammers can send thousands or even millions of spam emails at negligible cost, the recipient pays a considerable price for receiving this unwanted mail. Decreases in worker productivity, available bandwidth, data storage, and mail server efficiency are among the major problems caused by the reception of spam. This paper presents the technical survey of the approaches currently used to handle spam.

1. Introduction

Spam is unsolicited commercial email sent in bulk; it is considered an intrusive transmission. These bulk messages often advertise commercial products, but sometimes contain fraudulent offers and incentives.

Due to the nature of Internet mail, spammers can flood the net with thousands or even millions of unwanted messages at negligible cost to themselves; the actual cost is distributed among the maintainers and users of the net. Their methods are sometimes devious and unlawful and are designed to transmit the maximum number of messages at the least possible cost to them. Unfortunately, these emails impose a significant burden upon recipients [1]. Due to the dramatic increase in the volume of spam over the past year, many email users are searching for solutions to this growing problem. This paper presents a technical survey of approaches currently used to handle spam.

1.1 Problems encountered due to spam

The huge amount of unwanted email has led to significant decreases in worker productivity, network throughput, data storage space, and mail server efficiency. In large organizations, a considerable portion of the time of each worker is spent reviewing and deleting the spam itself, leading to a decrease in productivity. The increased network traffic has a deleterious effect on network performance, in general, and on the organization's mail server(s), in particular. Also, data storage space is consumed by the need to store the large volume of mail.

1.2 Why is Spam an issue?

Spam has been seen for quite some time now and could be considered the junk mail of the 21st century. It is growing at an alarming rate. The percentage of emails that are spam appears to have quadrupled between 2001 and 2003 and now accounts for somewhere around 40% of all email. It is also expected that that spam will increase to be over 60% of all email sent by the end of 2004. It can be advertisements for low mortgage rates or sales on the latest electronic devices. It can be offensive like advertisements for drugs or pornographic websites. It can also be hostile and contain viruses, Trojan Horses, or other malware.

In the case of spam it is more of a nuisance, but in sufficient volume it can present problems affecting productivity, bandwidth, and storage.

It is clear that as spam rises, the value of email as a business tool within corporate institutions will diminish. Assuming 10% of total mail is spam, and each employee spends 20 seconds/day deleting that spam, the estimated annual cost of spam to 10000-person company is \$675,000.

This is assuming that only 10% of email received is spam. If you bump this number up to 40%, the costs involved also shoots up. We should keep in mind that these costs are purely derived from lost productivity and do not include the costs to increase storage capacity nor the need to purchase more bandwidth to keep network traffic flowing.

The offensive spam may affect different people in different ways. Some may ignore it, while others may be deeply offended by it. Employers can be held liable when an employee sues based on a hostile work environment, if the company was aware of the issue and has not acted on it. Since spam originates from outside of your company, it is considered as a vendor or client harassing one of your employees. If you are aware of it, then it is your responsibility to take steps to remove it.

Employers face serious penalties if they don't remove such things from the working environment. People who have been subjected to harmful work settings can sue for up to \$300,000 in compensatory and punitive damages provided the company has more than 500 employees. Damages are scaled back to \$200,000 if the company has between 200 and 500 employees. If an employee leaves because of an environment judged hostile, they can ask for reinstatement, back pay and back benefits. [30].

In my opinion, the most significant risk is that spam would be considered hostile. These messages may contain viruses, Trojan Horses, worms, and web bugs among other

things. The senders may try to fool recipients into believing that the email is safe and is from a trusted source by using the names from the address book. Without proper precautions in place (virus and spam protection), this malware can spread like wildfire in an enterprise environment and bring messaging and network infrastructure to its knees. For Example, consider a Microsoft Exchange messaging infrastructure with 5000 employees. Introduce one worm on one of the workstation and it begins sending itself to all 5000 employees in the global address list. A few more worms get installed on other workstations and start replicating in the same manner. In a very short time the messaging load can clog messaging queues and network segments leading to slow network response or DOS (Denial of Service). Server storage space may be depleted resulting in legitimate email being lost or returned as undeliverable.

1.3 Background

The origins of spam can be traced to 1997, according to a commentary by Todd Burgess [2]. Even though a few reports and articles were published, the topic did not attract much attention at that time. Today, spam has turned out to be a nightmare for many in world of email and online electronic messaging services.

According to the report “Spam: 2003 Progress report,” [3] spam has increased by slightly over 200% during the period from March 2002 to March 2003. If this trend continues, electronic mail may become useless in a few years. Many nations, including the United States and various individual states, have begun implementing ant-spam laws to deter this practice; unfortunately, spammers have proven to be adept at modifying their techniques to escape detection.

2. Anti-Spam Techniques / Approaches

Anti-spam methods can be grouped into a few, fairly well defined, categories, though only some of these methods are currently in use.

There are two aspects to the response to spam. The most commonly discussed problem relates to the ability to distinguish between spam and legitimate email. For a large percentage of email, the decision is easy. We can easily identify more than half of

all email as either definitely legitimate (white) or definitely spam (black). It is the rest that is the most difficult to handle. We call these mails as “gray mail”.

The second issue for any comprehensive spam solution is the proper response to black and gray email. For confirmed spam, the solution is often easy simply delete them. However, there may be instances where other or additional actions are appropriate. Some possibilities are:

1. Forward the spam to the abuse department at the domain of the originator.
2. Reply to the originator voicing displeasure at receiving spam.
3. Reply to the originator to advise them that the email was not delivered.
4. Report the spam to a spam gathering station.

This list is not exhaustive and multiple responses may be appropriate in some situations. For gray mail, the appropriate response is unclear. Thus, the goal of a comprehensive anti-spam product is to be able to identify every email as either white or black with a very high probability of accuracy.

2.1 Blacklists and whitelists

Over time, patterns form in the receipt of email and a large percentage of the user’s legitimate email comes from a stable set of correspondents. Whitelists leverage this fact by allowing users to specify legitimate correspondents in a file that is used to screen incoming mail. All email from these “good guys” is delivered without further filtering.

Conversely, when mail from a particular email address is identified as spam, it is unlikely that any useful email will ever originate from that address. These “bad guys” addresses are placed on a blacklist and all email from blacklisted addresses is deleted without further evaluation. Occasionally, an entire domain may be identified as containing all bad guys and email from that entire domain is blacklisted. Sometimes blacklisted mails are stored in a distinct folder so that the recipient can later inspect them to ensure that no legitimate email is missed.

Using whitelists and blacklists is not without difficulties. Spammers have been able to avoid detection by spoofing addresses (Impersonating other users) or simply by changing their user names or IP addresses. More importantly, there is also a problem of mail from unknown sources, which cannot be put in the blacklist or whitelist. Overall,

blacklists and whitelists tend to stop 5-10% of spam [1]. Most of the tools available in the market use these lists as primary tools.

Some existing applications using blacklists in the market are listed below [18]:

- RBL (Real-time Blackhole List)
- MAPS RSS (Relay Spam Stopper)
- SBL (Spamhaus Block List)

2.2 Heuristic Engines

Keyword based systems are one of the most effective means of classifying email as spam or not. Heuristic engines operate by keyword filtering. They rely on a set of rules engineered by humans that is used to distinguish spam from legitimate email. They search for catch phrases, which are the most frequently repeated words in any spam-like email. Examples of catch phrases are “Get Rich” and “Free Viagra”. A scoring system or a point-value system is also employed to indicate the “spamminess” or likelihood that particular mail is spam according to the rule defined.

Even though heuristic engines are adaptive, they can be defeated by cleverly modifying mail to prevent detection. Thus, it provides better results when combined with other methods, which provide additional checks.

2.3 Authenticated Email

A subtle assumption about spam is that, to continue to be effective, spammers must protect their real identities. Authenticated email utilizes this fact to filter out email that cannot be strongly attributed to a known or otherwise viable entity. At its simplest level, authenticated email augments a whitelist approach by ensuring that messages are authentic; that is, determining that the originating address was not spoofed or impersonated.

Authenticated email is unlike whitelists in that very little a priori knowledge is assumed. Rather, they rely on intuition about conflicting properties of legitimate email and spam. For example, email originators give thought and effort to each email and most legitimate email is drafted for a small number of recipients. Unfortunately, spammers go to great lengths to disguise the number of recipients and it is not always

easy to tell if a message is sent to one, a few, or many recipients by reviewing the email text or headers.

One method of determining that an email is not spam is to guarantee that the originator committed a significant amount of effort to delivery of the message. Presenting a challenge in the form of a mathematical computation or pictographic puzzle can ensure this effort.

Pictographic puzzles are designed in such a way that only humans can compute the proper reply. Though it takes only a few seconds per email, it would be impossible for a spammer to personally solve enough of these puzzles to be worth his or her time. Other puzzle mechanisms rely on computationally intensive problems that do not require input by a human. The foundation of these techniques is that the spammer computer could not solve enough of these puzzles to be worth the time committed since automated spam engines are slowed drastically by puzzles.

Both of these techniques rely on email proxy system and operate by making email delivery have a small but nontrivial cost to the originator. Even though this might be slightly inconvenient to legitimate users wishing to communicate with a genuine cause, it helps to prevent spam to such an extent that it has been employed very widely in products available in the market today.

The puzzle method has its drawbacks, however. It may lead to frustration among legitimate senders and they may sometimes choose not to send any mail instead of wasting their time by responding to the challenges. It is also possible that, under a high volume of spam, the proxy will become an email bottleneck. This may lead to a disruption of the activities of the recipient's server itself.

2.4 Distributed Checksum Clearinghouses

Distributed Checksum Clearinghouses gather known spam emails and store their patterns (fingerprints) in databases that can be used by anti-spam systems. There are two major ways this method can be used. First, they can match an email pattern with a particular fingerprint stored in a database. This serves to identify mail with patterns identical to known spam fingerprints. Second, they can use more complex fingerprints generated via the analysis of many emails (similar to the methods used in the detection of computer viruses). This method compares the fingerprints to the new

mails to find any matching patterns. They can also be trained to tackle and identify randomization of spam, which is the insertion of random text in the spam to escape detection.

2.5 Rule-based Ranking/Scoring

This method compares a new message against a large number of stored spam patterns. The patterns are usually in the form of regular expressions and are used to compute a numerical score for each email. When the pattern of an incoming mail matches an existing pattern, its score is increased. If a message's score exceeds a given threshold value, it is labeled as spam; otherwise, it is taken to be a legitimate mail.

The current ranking/scoring rules are based on current spamming techniques and involve search for phrases like "Herbal Viagra" or "heirs of African dictators". They will utilize other phrases in the future to keep up with changes in common spamming topics. This creation of new rules, involving new computations, increases the effectiveness of this method with time.

Rule-based ranking is one of the latest anti-spam methods and is proving to be highly effective. The most popular tool employing this is *SpamAssassin*.

2.6 Distributed Blacklists

A distributed blacklist is a network tool for anti-spam engines. It is a compilation of known spammer email addresses and domains acquired from various sources.

2.7 Honeypots

Honeypots are dummy email addresses that are created to attract spam. They list the known instances of spam in a database that can be accessed by other potential recipients of the same spam and used to block delivery. The problem encountered in the usage of this method is similar to the one in the heuristic engines. They help in filtering out known spam, but they cannot help block previously unknown spam.

2.8 Reverse DNS Lookup

When an email is sent from one server to another, a TCP/IP connection is made between the two servers. The mail server that is receiving the email can take the IP address of the sending server and do a DNS lookup on that address to see if it matches what is in the header information of the email. This is a means of finding out if the sender is attempting to spoof the address from where the mail is actually originating.

2.8 Statistical Classification Engines

The first generation of spam filters used rules to recognize specific spam features. Now a new generation of statistical spam filters seems to offer significantly better performance. Statistical filters look at the entire contents of each incoming email and decide whether it's spam based on its overall similarity to previous spams. This new kind of filter routinely catches over 99% of current spam with near zero false positives.

One of the Statistical Classification Engine is the Bayesian filter. Bayesian filters are the latest in spam filtering technology. They recognize spam by looking at the words (or 'tokens') they contain.

A Bayesian filter starts with two collections of mail, one of spam, and one of legitimate mail. For every word in these mails, it calculates a spam probability based on the proportion of spam occurrences. For example, "Guaranteed" has a spam probability of 98%, because it occurs mostly in spam; "This" has a spam probability of 43%, because it occurs about equally in spam and legitimate mail; and "deduce" has a spam probability of only 3%, because it occurs mostly in legitimate mail.

When a new mail arrives, the filter collects the 15 or 20 words whose spam probabilities are furthest (in either direction) from a neutral 50%, and calculates from these an overall probability that the email is spam.

Because they learn to distinguish spam from legitimate mail by looking at the actual mail sent to each user, Bayesian filters are extremely accurate, and adapt automatically as spam evolves.

The simplest statistical filter can be described in a paragraph. Users discard all their spam in a separate trash can. At intervals, a program looks through all the user's email and, for each token, calculates the ratio of spam occurrences to total occurrences. For example, if "cash" occurs in 200 of 1000 spams and 3 of 500

nospam emails, its spam probability is $(200/1000) / (3/500 + 200/1000) = 0.971$. When a new email arrives, extract all the tokens and find the fifteen with probabilities $p_1 \dots p_{15}$ furthest (in either direction) from .5. The probability that the mail is a spam is

$$\frac{p_1 p_2 \dots p_{15}}{p_1 p_2 \dots p_{15} + (1 - p_1)(1 - p_2) \dots (1 - p_{15})}$$

These statistical filters have some important benefits:

- 1) **They are very effective.** Even the simplest statistical filter will catch 99% of current spam. The most effective filter available now is Bill Yeraunus' CRM114, catches 99.8% of spam.
- 2) **They generate few False Positives.** False positives, legitimate emails that are mistakenly treated as spam, are the bane of spam filtering. Statistical filters yield fewer false positives because they consider evidence of innocence as well as evidence of guilt. A token that occurs disproportionately often in your nospam mail, like the name of a friend, will count as much toward decreasing the spam probability as a token like "cash" would to increasing it.
- 3) **They Learn.** You don't have to look through piles of spam and figure out rules to identify them. Whatever's in there, the filters tend to find it. Like us, statistical filters notice that the token "cash" is sign of spam. However, they also notice that "modalities" (used in a surprisingly high proportion of Nigerian spams) and "FF0000" (html for bright red) are even better signs of spam. And as spammers change their messages or their infrastructure, the filters adapt.
- 4) **They let each user define what's spam.** Although statistical filters could be used at the network level, ideally the probabilities should be calculated individually for each user. To the extent users' definitions of spam differ, their inboxes will reflect this.
- 5) **They are hard to trick.** There are only two ways to get past a statistical filter: use fewer bad words, or use more innocent words. Spammers can't do the latter, because the most innocent words (words related to your friends and family, your work, your interests) vary for each user. So they have to use fewer bad words. They can't use weird spellings (e.g. "Freee" instead of "Free") because filters

quickly learn those. Their only option is to use vaguer and vaguer euphemisms, or simply to have some generic sounding text, and a link.

Spammers also try to prevent filters from recognizing the tokens in the mail by breaking them up-- for example, by using white space or punctuation characters in the middle of words

Li ke th.is

But this doesn' t work well either. One reason is that legitimate email doesn' t have many individual letters or word fragments in it, so a fragment like "ke" or "th" will tend to have an above-average spam probability. Another is that they can' t do this sort of obfuscation on headers and urls, and those are enough by themselves to identify most spam. We could probably reconstruct the broken words if we had to, but this hasn' t even been necessary so far.

Spammers sometimes insert html comments at random places within words, but this is also easy to ignore. In general, on the token front, it is a question of closing loopholes. There are only so many tricks spammers can use, and we deal with them individually. So far none has been insurmountable.

People sometimes ask, what if spammers sent the mail as an image? They do already, and this kind of spam is easy for filters to catch. Tokens like "img" and "href" have spam probabilities like those of pornographic terms. Plus there is the domain name and filename in the url, and, as always, the headers. On the whole, spam containing html is easy to filter. The most hardened spammers seem to know this and already avoid html in their mails. Whatever the spam of the future looks like, it probably won' t contain html.

3. Common Methods for Harvesting Email Addresses

Spammers use various methods to gather valid email addresses. Some of the methods by which spammers harvest email addresses are listed below. If you follow the recommendations below, you can dramatically decrease the amount of unwanted email sent to you.

- **Extracting mail addresses from mailing lists and directories:** The easiest method of obtaining valid email addresses in bulk is through the mailing lists and directories (white pages or yellow pages) available online. The spammers buy mailings lists or

they hack into websites hosting such lists. The mailing lists of huge corporations are an easy target. Computer “robot s” allow spammers to gather email from online directories. Membership into groups and discussion forums also provides access to the spammers to acquire valid email addresses in a large number.

Recommendation 1: Even if sites promise to protect your email address, assume it can be obtained anyway. When possible, don’t post your email address in any directory. If you must provide an email address, use a disposable email address when registering with a site or buying a product.

Recommendation 2: In the case of reputable companies, alter your personal preference options to specify that your contact information should not be shared with others.

- **Harvesting email addresses listed on web pages:** The spammers also employ programs which weave through web pages in search of valid email addresses. They collect valid email addresses and transmit them back to the spammer.

Recommendation 1: Use an Internet search tool like Google.com to find all occurrences of your email address on Internet. Then go to each of those sites and request that the Webmaster remove your email address from that webpage (specify the URL).

Recommendation 2: Protect email addresses when you place them on a webpage. They can be made human interpretable. For example, if your email address is *marysmith@aol.com* write, "marysmith at aol dot com".

- **Forms filled out on paper and on the web:** Some companies get hold of the addresses of all the users who fill out forms on paper on the web for them. These addresses are sometimes sold to spammers.

Recommendation 1: When filling out web surveys and registration forms do not give out your email addresses. Look for a check box that asks you if it is okay to send similar offers or information to you. Be alert for options that highlight themselves automatically to include your email addresses for further communication from partner and related sites.

Recommendation 2: Do not give out friends’ email addresses to any service. There are services that say “refer a friend” for bonus. These may be simply email address

harvesting services that get two mail addresses at one shot. For example, there are several free greeting card sites that ask for a friend's email address to gain a bonus, but they turned out to be making money out of the email addresses they gathered.

Recommendation 3: Use multiple email addresses for different purposes. You can also use "disposable email addresses", which are used to consolidate various addresses but allow you to shut off any address, which is attracting spam. Many sites provide free email address and one-time email addresses.

- **From IRC and Chat rooms:** People in the chat rooms are often willing to give their addresses anyone who asks, making the work of a spammer even easier. People new to net activities are often easy targets.

Recommendation: Don't use your email address as your chat id and don't give out your email address in public "chats".

Recommendation: For America Online and similar Instant messenger services, remove your online profile information.

- **By guessing and cleaning:** The spammers sometimes send messages by guessing a particular address from a user's first name and last name and wait for a confirmation or an error message to return from that address.

Recommendation 1: Make email addresses long but not incomprehensible. Shorter email addresses are easy to guess by brute force attacks and dictionary attacks.

Recommendation 2: Do not reply to any junk email directly, even to tell instruct them to remove your address from their email address. This action will confirm the existence of a valid email address to the spammer. Even if the spammer may not use your email address to send spam directly, he may sell your address to other spammers

Recommendation 3: Set your email Inbox to operate with the option of the preview pane closed to avoid confirming to the spammers that you have opened and viewed their mail.

Recommendation 4: Set your email client to "never" confirm receipt of email.

- **Using social engineering or deception:** It is often surprisingly easy to trick people into giving out valid email addresses by impersonating system administrators or offering bogus giveaways or contests.

Recommendation 1: Do not sign up for services, which announce “Free drawings” and “Lottos”. Many of them collect email addresses that are valid and send junk email.

Recommendation 2: Do not purchase anything from suspect parties, even if the product looks very attractive. Spammers fake their email addresses and spoof their source address to escape detection and then send thousands of bogus email messages.

Recommendation 3: Do your best to avoid opening suspicious mail. This can be done by looking at the subject line, originating address, etc Simply delete the suspected mail or, open it when your internet connection is closed to prevent any hidden bugs from working for the spammer.

Recommendation 4: Avoid using the “Reply All” option when you receive email from someone outside your organization. This may give many addresses to the spammer in the reply. This problem can be tackled by typing in the additional email addresses in the “Bcc” field instead. Using the “Bcc” hides the other recipients from viewing all the email addresses to which the message is being sent.

- **From domain contact points:** Some domains usually have contact points (such as administration, technical, or billing) which also list the addresses of persons related to that contact point. This provides direct contact information.

Recommendation 1: Use generic email addresses for contact points such as customerservice@yourcompany.com.

Recommendation 2: Report spam to abuse at the particular domain when you receive a message via their email service. For example, when a junk mail is received via yahoo mail, sending them a mail addressed to abuse@yahoomail.com helps the yahoo people to avoid further communication from the spammer.

- **Scanning newsgroups for email addresses:** A common method of acquiring email addresses is scanning newsgroups (UseNet) for valid user addresses. Spammers can also scan the headers and bodies of available mail and check for the occurrence of symbol ‘@’ to find valid user addresses.

Recommendation 1: Post anonymously and protect your email address from newsgroups when possible.

Recommendation 2: When joining a newsgroup or a forum, make sure that the messages/replies do not reach your email address. Instead, make them post the messages in the forum itself so that your address cannot be traced.

4. Evaluating the cost of Spam

4.1 Overall Cost of Spam

According to Ferris Research, the total cost of spam to corporate organizations in the United States in the year 2003 was \$8.9 billion. The Ferris estimate is based on 3 spam messages per day for the average user and 20 spam messages a day for the highly exposed user, with 4.4 seconds to take action against the message.

4.2 Calculating Productivity Loss

Lost productivity results from users spending time sorting through their Inbox to weed out spam messages from legitimate messages. On average, users spend 4.4 seconds per spam message to determine if the message is indeed spam, and then take action. Some advanced users can identify and delete spam in bulk, but they also have a higher risk of accidentally removing legitimate messages. Other users take much longer to remove spam, but are at less risk of losing legitimate email.

An example of the spam cost worksheet:

Number of employees with email 500
Average annual salary \$60,000
Average spam per day per employee 25
Seconds to identify and delete each spam 4.4

Total salary lost daily \$458.33
Total salary lost monthly \$8,975.69
Total salary lost annually \$107,708.33

Total time lost daily 15.28 hours
Total time lost monthly 37.40 work days
Total time lost annually 448 work days

* based on a 220 day work year

4.3 Legal Liability

Allowing offensive material into the work place can be a substantial liability. Whether it is by creating a hostile work environment, or the perception of prejudice or lack of sensitivity, a single offensive spam message could be very expensive. On average, a sexual harassment claim based on an offensive spam costs between \$72,000 and \$500,000 per incident. Human Resource departments are reporting a record level of written claims of harassment related to offensive, sexually implicit, unsolicited commercial email.

Spam is more than simply a network resource drain. David Woodall, the head of information technology for CIO Magazine, says, "Now, people are saying they feel harassed by it. It' s gone from a technical issue to a human resources issue." In a recent survey by Strategic Surveys International of Fortune 500 companies, Chevron Corporation and Morgan Stanley Dean Witter have both settled multimillion-dollar sexual harassment lawsuits as a result of internally circulated emails that contained offensive content.

With available anti-spam technology, an employer can take measures to protect employees from unsolicited, potentially offensive material as part of providing a nonhostile or non-offensive work environment. For the employer, anti-spam technology provides a measure of protection against potentially expensive legal liability.

4.4 Resource Consumption

IT resource consumption costs include not only network bandwidth and disk storage, but also the cost of dealing with spam related inquiries. As the level of spam increases, users become increasingly annoyed, and complaints to the help desk increase. Some complaints concern specific offensive messages, while other complaints concern the overall volume of daily spam. Depending on interruption time, the average cost of a help desk call can be from \$15 per incident to as much as \$35 per incident.

Additional infrastructure resources, such as network bandwidth, disk storage, and message store processing, are also heavily impacted by spam. If spam in your organization represents 40% of all incoming messages, that translates to 40% more processing and storage capacity that your email system will be required to sustain. By

eliminating spam, and thereby increasing network bandwidth, disk storage and email system processing, the email system infrastructure will regain the lost resources consumed by spam.

Bibliography

- [1] "What is spam?" <http://spam.abuse.net/overview/whatisspam.shtml>
- [2] "Fighting Email Spammers: A Commentary", Todd Burgess
<http://eddie.cis.uogelph.ca/~tburgess/local/spam2.html>
- [3] "Spam: 2003 Progress Report", ZDNet,
<http://webevents.broadcast.com/ZDWebcast/spammer/index.asp?loc=11>
- [4] White Paper: Spam Classification Techniques, mxlogic.
- [5] "Spam Filtering Techniques – Comparing Half-Dozen Approaches to Eliminating Unwanted Mail" <http://gnosis.cx/publish/programming/filtering-spam.html>
- [6] "Email Harvesting Techniques"
http://secinf.net/anti_spam/Email_Harvesting_Techniques_FAQ.html
- [7] Network World Fusion Anti-Spam Products.
www.nwfusion.com/research/2002/0513spamside3.html
- [8] Trend Micro Product. Interscan eManager.
<http://www.trendmicro.com/en/products/gateway/isem/evaluate/features.htm>
- [9] Mailwasher product. www.mailwasher.net
- [10] Gordano Anti-Spam Technology, www.ntmail.co.uk/technology/anti-spam.htm
- [11] Lyris Mailshield Server,
http://www.lyris.com/products/mailshield/server/how_it_works.html
- [12] GFi MailEssential's Product, <http://www.gfi.com/mes/index.html>
- [13] ContactPlus Products SpamBuster, www.contactplus.com/products/spam/spam.htm
- [14] GBS Design Inc. Inbox Protector, www.gbs-design.com/InboxProtector
- [15] CloudMark Product- Spamnet,
<http://www.cloudmark.com/products/spamnet/learnmore/howitworks.php>
- [16] About Email Spam Products, <http://email.about.com/cs/macspamreviews/>
- [17] Dignity Software product Spam Alarm,
<http://www.dignitysoftware.com/spamalarm.php>
- [18] <http://email.about.com/cs/winspamreviews/tp/antispam.htm>

- [19] SurfControl's Anti -Spam Agent,
http://www.surfcontrol.com/products/content/internet_databases/antispam/
- [20] SpamCop Email system for individuals, <http://mail.spamcop.net/individuals.php>
- [21] SpamAssassin product, <http://au.spamassassin.org/index.html>
- [22] Pyzor Spam-catcher, <http://pyzor.sourceforge.net/>
- [23] Digiportal Software-ChoiceMail, <http://www.digiportal.com/>
- [24] Frontbridge Spam Filtering, <http://frontbridge.com/spam-filtering/>
- [25] <http://www.sourceforge.net>
- [26] <http://email.about.com/cs/winspamreviews/gr/spamihilator.htm>
- [27] http://www.bearcave.com/software/mail_filter/index.html
- [28] Natural- Born Spam Killers, Daniel Tynan, From the May 2003 issue of PC World magazine, <http://www.pcworld.com/resource/printable/article/0,aid,109698,00.asp>
- [29] Real Time Stats. <http://www.appriver.com/>
- [30] McCullagh, Declan. CNET News.com. 'Porn Spam – Legal minefield for employers. April 7, 2003. URL: <http://rss.com.com/2100-1032-9957658.html>