



**Institute for Internet Security**  
University of Applied Sciences  
Gelsenkirchen

Neidenburger Str. 43  
45877 Gelsenkirchen  
Germany

# Anti-spam measures of European ISPs/ESPs

A survey based analysis of state-of-the-art  
technologies, current spam trends and  
recommendations for future-oriented  
anti-spam concepts

Christian Rossow  
August 2007

***Anti-spam is like a pet.***

***Some people have great pets, through hard work at training, respectful treatment, and exercise. Other people expect the critters to look after themselves and complain when there's crap on the carpet.***

*Posted by "Lee" on LinuxSecurity.com*

## Anti-spam measures of European ISPs/ESPs

<b>1</b>	<b>Preface</b>	<b>1</b>
1.1	Introduction	1
1.2	Scope of this document	1
<b>2</b>	<b>Introduction</b>	<b>2</b>
2.1	Description of spam	2
• 2.1.1	Definition(s) of spam	2
• 2.1.2	Extent of abuse email	3
2.2	Infrastructure of ISPs/ESPs within Europe	4
• 2.2.1	Differentiating ISP and ESP	4
• 2.2.2	Internet penetration in Europe	5
• 2.2.3	ISPs/ESPs landscape in Europe	5
• 2.2.4	Components of an email system	6
<b>3</b>	<b>Current trends of spam</b>	<b>8</b>
3.1	Spam via botnets	8
3.2	Localisation of spam	9
3.3	Image spam	9
3.4	Domain tasting	10
3.5	Stock spam	11
<b>4</b>	<b>Categorising anti-spam methods</b>	<b>12</b>
4.1	Basics of SMTP	12
4.2	Architecture of anti-spam methods	13
4.3	Different efficiencies of anti-spam methods	15
<b>5</b>	<b>Overview of anti-spam methods</b>	<b>16</b>
5.1	Email envelope analysis	17
• 5.1.1	Blacklisting	17
• 5.1.2	Greylisting	18
• 5.1.3	Whitelisting	19
• 5.1.4	Sender authentication	19
• 5.1.5	Sender address verification (SAV)	22

---

5.2	Email data analysis .....	24
•	5.2.1 Heuristical methods .....	24
•	5.2.2 Statistical methods .....	25
•	5.2.3 Blacklisting of URIs (URIDNSBL) .....	26
•	5.2.4 Checksum comparison .....	27
5.3	Structural adjustments .....	28
•	5.3.1 Splitting message submission from message relay .....	28
•	5.3.2 Proof-of-work .....	29
•	5.3.3 Challenge-response mechanisms .....	30
•	5.3.4 Electronic postage .....	31
•	5.3.5 Traffic shaping .....	31
5.4	Useful anti-spam tools .....	32
•	5.4.1 Spamtraps .....	32
•	5.4.2 Reputation systems .....	33
•	5.4.3 Frequency analysis .....	34
<b>6</b>	<b>ENISA survey on anti-spam</b> .....	<b>36</b>
6.1	Description of the survey .....	36
6.2	Discussions of the results .....	37
•	6.2.1 Spam is a critical security threat .....	37
•	6.2.2 Analysing spam .....	37
•	6.2.3 Most used spam-filtering measures .....	38
•	6.2.4 Efficiency of anti-spam methods .....	38
•	6.2.5 Review of sender authentication .....	39
•	6.2.6 Protection against outgoing spam .....	40
<b>7</b>	<b>Empirical research on blacklisting</b> .....	<b>41</b>
7.1	Description of research methods .....	41
7.2	Origin of data .....	41
7.3	Assessment of blacklists .....	43
7.4	Coverage of blacklists .....	45
•	7.4.1 Reputation of the IP address space .....	45
•	7.4.2 Status of each blacklist .....	47
•	7.4.3 Potential for lists to block botnets .....	47

---

• 7.4.4 Recent vs. old spam sources .....	49
7.5 Blacklist entries by country .....	49
• 7.5.1 Regional biases of blacklists .....	49
• 7.5.2 Union of all blacklists .....	50
• 7.5.3 Discredited countries .....	50
7.6 Blacklist entries by AS .....	52
• 7.6.1 Discredited Autonomous Systems.....	52
7.7 Intersections between blacklists .....	53
7.8 Quality assurance for blacklists .....	55
• 7.8.1 Quality indicators for blacklists .....	55
• 7.8.2 Measuring false positive rates .....	55
7.9 Quality assurance for whitelists .....	57
<b>8 Anti-spam recommendations for European providers</b> .....	<b>58</b>
8.1 Highly recommended .....	58
• 8.1.1 Manage port 25 .....	58
• 8.1.2 Consider network level blocking .....	61
• 8.1.3 Support sender authentication .....	62
• 8.1.4 Offer user-defined anti-spam solutions.....	64
• 8.1.5 Contribute to anti-spam networks.....	64
8.2 Controversial anti-spam methods .....	65
• 8.2.1 The future of email data analysis.....	65
• 8.2.2 Latent damage through SAV .....	66
• 8.2.3 (R)evolution of structural adjustments .....	67
<b>9 Conclusion</b> .....	<b>69</b>
<b>A Glossary</b> .....	<b>71</b>
<b>B References</b> .....	<b>76</b>
<b>C Annex - ENISA survey</b> .....	<b>77</b>
<b>D Annex – IP address based blacklist entries by country</b> .....	<b>81</b>
<b>E Annex - Blacklist entries by AS</b> .....	<b>93</b>
<b>F Annex - Graphical blacklist coverage</b> .....	<b>105</b>

# **1 Preface**

## **1.1 Introduction**

This document arose during a traineeship at ENISA, the European Network and Information Security Agency, from June until August 2007. It is the bachelor thesis of the author to complete his studies of Applied Computer Sciences at the University of Applied Sciences Gelsenkirchen in Germany. This document is especially designed for network operators working for providers and dealing with electronic communication via email. This document can be considered as a common paper on anti-spam measures.

Nowadays spam is an old topic, but still it is a rising problem. Spammers manage to circumvent current anti-spam installations and harm by consuming resources, damaging the reliability of email as a communication instrument and tricking recipients into reacting to spam. This document tries to review existing technologies, exposes current spam trends and gives recommendations to providers how to mitigate the spam problem.

The paper is based on data from a survey conducted by ENISA in June 2007, asking providers of the European Union for their anti-spam best practices and future implementations. Three of the top-10 European providers participated in this study. All in all 28 answers have been received. Due to the locality of ENISA this document is based on data likely to be from the European Union (EU), though most of it applies for non-EU providers.

A special focus of this document is IP level blacklisting. Some research was made in order to estimate the usefulness of blacklisting and to show how the most well-known blacklists work. Moreover regional blacklist statistics and association between list entries and providers were made.

This document is a comprehensive guideline for providers, giving new aspects based on the empirical research on blacklists and the survey conducted by ENISA. It will help to improve the efficiency of most providers' anti-spam installations and vice versa support providers to get a better, non-abusive reputation.

## **1.2 Scope of this document**

This document concentrates on anti-spam mechanisms for email communication. Although some methods can be applied for other kinds of spam like unsolicited VoIP telephony or instant messaging spam, all following contents are specialised on email spam.

Moreover this document does not cover legal aspects of spam. National legislation on spam is quite different from country to country, and therefore (if considered) only laws from the European Union are covered. For more information on legislation each provider should contact appropriate persons like lawyers or national regulation authorities. However, in some parts of the documents hints are given than can be applied to common national laws.

This document does not review special anti-spam products, but instead describes methods or tools generally. It is not considered important to give specific tools for the described anti-spam installations. Thus the main part describes the methods themselves and gives pros and cons rather than an evaluation of individual products.

## 2 Introduction

### 2.1 Description of spam

#### 2.1.1 Definition(s) of spam

Spam is a term, which has several definitions. It depends on the attitude, position, temper or even mood of the person receiving an email, whether he/she classifies an incoming mail as spam. On the one hand people can be very strict and classify each email they do not want to read as spam. This could also include the daily fun emails sent by colleagues, newsletters that are not (or not anymore) desired as well as information from suppliers about new services. On the contrary, other people might want to receive these messages and would not classify them as spam.

Every anti-spam method has to deal with this big problem. It is an organisational question, which should be answered before getting technically. Existing definitions help to limit this problem. Often referenced definitions of spams are:

*“Spamming is the abuse of electronic messaging systems to send unsolicited bulk messages, which are generally undesired.”*

*Wikipedia.org*

*“An electronic message is spam if:*

- (1) the recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients AND*
- (2) the recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent.”*

*Spamhaus.org*

Some parties even avoid using the term spam because of the varying definitions of spam, depending on the local legislation of the country.

For this document spam will not be re-defined, but is understood as a shortcut for unsolicited commercial email. In order to consider legal aspects as well, the EU Directive 2002/58/EC on Privacy and Electronic Communications (October 2003)<sup>1</sup> gives an overview about the juristic framework, which all EU member states must implement into national legislations. Within this document commercial email communication is restricted:

*“The use of (...) electronic mail for the purposes of direct marketing may only be allowed in respect of subscribers who have given their prior consent.”*

*EU Directive 2002/58/EC Article 13 §1*

<sup>1</sup> The EU commission published the Directive 2002/58/EC on the web: [http://eur-lex.europa.eu/pri/en/oj/dat/2002/l\\_201/l\\_20120020731en00370047.pdf](http://eur-lex.europa.eu/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf)

This method is commonly known as opt-in mechanism. On the contrary the opt-out mechanism does not require the receiver to permit the communication beforehand. The latter is not allowed in the EU member states. But in practice there are varying interpretations about the EU Directive 2002/58/EC. This leads to similar, but not equal legislations regarding spam within the EU<sup>2</sup>. Furthermore a large part of spam originates in Non-EU countries that sometimes have no or very weak legislation on spam.

Moreover the opt-in method is a bit softened through the “soft opt-in” mechanism, which allows email communication due to existing business relationships:

*“Where a natural or legal person obtains from its customers their electronic contact details for electronic mail (...), the same natural or legal person may use these electronic contact details for direct marketing of its own similar products or services provided that customers clearly and distinctly are given the opportunity to object (...) such use of electronic contact details when they are collected and on the occasion of each message (...).”*

*EU Directive 2002/58/EC Article 13 §2.*

Due to the fact that the term ham is often used in this paper a definition of it must be given:

*Every email not considered as spam is ham, i.e. ham = all emails - spam.*

### 2.1.2 Extent of abuse email

European Internet users are weary of receiving spam. They are stressed out of reading unsolicited emails, often with non-ethical or unserious content. While a fifth of all users do not care about spam, almost a half of all users would consider switching the provider if they had too much trouble with spam or viruses<sup>3</sup>. This alarming figure should motivate providers to improve their spam- and virus mitigation techniques.

Spam has been an increasing problem world-wide since the beginning of email communications. Many public studies provide data about spam. Some of them even try to figure out the arising costs. It depends on the extent, the point of view, the locality and the goal of the studies, which data is collected and published.

ENISA studied the spam problem within the EU and neighbouring countries by the bias of two surveys in 2006. The questionnaire has been sent to technicians in order to make out the dimension of spam. Goal of the surveys among other things was becoming aware of a security/spam problem.

Based on an evaluation of approximately 490 million mailboxes the internationally active MAAWG installed a metrics and reports data about spam quarterly<sup>4</sup>. Current data of 2006 and the first quarter of 2007 have shown that spam is getting a very important security issue. The report documents an abusive email rate of 75-80%, the estimated number of

<sup>2</sup> Referring to the results of an ENISA survey conducted in 2006 still 34% of the participants responded that direct marketing not necessarily needs an opt-in confirmation by the recipient, among this anyhow 59% EU member states. See section 3.3.4 of the [ENISA1] for details.

<sup>3</sup> See [EStat01] chapter 3.2.6 for more details.

<sup>4</sup> See <http://www.maawg.org/about/EMR> for detailed reports.



unknown cases may in fact be higher. Comparable results were found by MessageLabs Intelligence in their security report 2006<sup>5</sup>.

It is difficult to determine the precise rate of spam, because the measures on the network level drop email submissions before accepting data. An exact forecast of the amount of emails, which might have been delivered through a dropped connection is not possible. Assuming that each connection leads to at least one email is possible, though, and gives rough ratios.

Another danger is the rising ratio of 35% image spam<sup>6</sup> that causes a higher consumption of bandwidth on the network. Image spam has much higher traffic demands than text spam, stressing the connections of providers. Commtouch's report about spam in 2006 documented a ratio of 70% overall spam bandwidth taken by image spam, similar results were observed by SoftScan recently<sup>7</sup>.

Currently there is no end in sight. Growing botnets and more intelligent spam methods complicate the battle against spam. The email architecture runs into danger to loose confidence of the users. Phishing or stock spam hurt naïve users. Viruses infect more and more computers. Everyone is bothered by unsolicited messages and last but not least providers are worrying that lots of bandwidth is wasted by spam.

## **2.2 Infrastructure of ISPs/ESPs within Europe**

Service providers in Europe built up a huge market. Ten major organisations share more than half of the entire broadband market. This chapter describes two types of providers (ESP and ISP), gives an overview about the Internet penetration within Europe and gives an overview about the big players on the European provider market. Finally it describes the internal infrastructure of an email system used by these organisations.

### **2.2.1 Differentiating ISP and ESP**

The terms ISP and ESP are often used when talking about email communication and especially protection against spam. To explain the differences between both, definitions for these are given.

*An Email Service Provider (short: ESP) is an organisation offering email services to its subscribers. These services usually cover managing an own mailbox, i.e. receiving emails to and sending emails from it.*

*An Internet Service Provider (short: ISP) is an organisation offering access to the Internet to its subscribers. Additionally ISPs usually offer services of ESPs.*

<sup>5</sup> See [http://www.messagelabs.com/Threat\\_Watch/Intelligence\\_Reports](http://www.messagelabs.com/Threat_Watch/Intelligence_Reports) for detailed reports.

<sup>6</sup> Sophos observed an image spam rate of 35.1% within their annual security report of 2007, see <http://www.sophos.com/securityreport2007>; Commtouch observed a ratio of 35% image spam, see [Comm01].

<sup>7</sup> SoftScan published the study in March 2007 at <http://www.softscan.co.uk/composite-548.htm> and described the growth of the average file size of an email from 6.62kb to 11.76kb since September 2006. Combining this observation with the assumed image spam ratio of 35%, likely image spam bandwidth results can be computed.

Within this document it is assumed that an ISP generally offers email services. ISPs who do not offer email service will not benefit from this document. Simply put, both ISPs and ESPs offer email services, but only ISPs offer in addition a possibility to access the Internet.

### 2.2.2 Internet penetration in Europe

Two sources of information deliver data about the Internet penetration in Europe and the European Union. In July 2006 Eurostat figured out<sup>8</sup> that 40% of the households in the 25 EU member states had at least one access to the Internet, more than a half of these accesses are broadband connections.

Since these statistics only review private Internet use based on households, Internet World Stats<sup>9</sup> on the other hand does research on the total amount of Internet users. As of data from June 2007, it figured out that there are 256 million users within the EU and 322 million users in Europe. Europe is the continent with most Internet users (after Asia), covering more than a fourth of all Internet users.

More than half of all European Internet users live in only five states. Germany tops the list with 15.7% of European Internet users, followed by UK (11.7%), France (10.2%), Italy (9.8%) and Spain (6.1%). On the other hand smaller countries have a higher penetration rate, like Iceland with 86.3% of the population being online or Sweden, The Netherlands and Portugal all above 70%.

### 2.2.3 ISPs/ESPs landscape in Europe

The number of total providers in Europe is difficult to find. RIPE manages a list of members by country<sup>10</sup>, giving an approximate indication where providers act, grouped by the size of the provider. Since this table is more confusing than helpful when looking for the major providers in Europe, other sources have to be considered.

StrategyAnalytics publishes a quarterly report<sup>11</sup> of the 75 biggest European broadband service providers. Although this definition varies from the definitions of ISPs/ESPs, the figures are likely to be coherent with figures on the biggest ISPs/ESPs. Since the reports are only available for paid subscribers, some information from the public abstracts of these reports has been aggregated. In this way a fairly complete table of the top-10 providers was constructed.

The fourth place is probably taken by Telefonica (ES) and AOL Europe might be on the tenth place. For a complete list with accurate figures StrategyAnalytics fee required list should be considered.

Rank	Provider
#1	France Telecom (FR)
#2	Telecom Italias (IT)
#3	Deutsche Telekom (DE)
#4	?
#5	BT Retail (UK)
#6	Virgin Media (UK)
#7	Free (FR)
#8	United Internet (DE)
#9	Neuf Cegetel (FR)
#10	?

Table 1: Biggest European broadband service providers

<sup>8</sup> See [EStat01] chapter 3.2 for more information.

<sup>9</sup> See <http://www.internetworldstats.com> for more information.

<sup>10</sup> See <http://www.ripe.net/membership/indices/> for the members list.

<sup>11</sup> See <http://www.strategyanalytics.net/default.aspx?mod=ReportAbstractViewer&a0=3482> for the latest issue of the first quarter 2007, published end-June 2007.

Another interesting point of view is looking at the amount of spam geographically originating in Europe. Security vendor Sophos did research in this field and figured out<sup>12</sup> that Europe is the worst spamming continent with 35.1% of all spam originating in it. Five of the top 12 spamming countries – namely Poland, Italy, France, Germany and Spain – are from Europe. Similar frightening results were published by Spamhaus<sup>13</sup> with four European countries – namely UK, Germany, Netherlands and France – within the top 10 of spamming states. As a conclusion European Internet users threaten the world by sending spam, no matter which list is more accurate.

## 2.2.4 Components of an email system

It is necessary to specify an email system in more detail when talking about spam. Terms like “mail server” or “email protocol” are too general to be used in such a specific context. For this reason, all modules of a currently modern system are described here:

*A Mail User Agent (short: MUA) is a software used by a client in order to send and receive emails.*

*A Mail Submission Agent (short: MSA) is a process accepting email submissions from MUAs in order to forward them to recipients.*

*A Mail Transfer Agent (short: MTA) is a process for sending mails to and receiving mails from other MTAs, equally where located. Moreover it accepts email submissions from well-known MSAs to relay these mails.*

*A Mail Delivery Agent (short: MDA) is a process used by the recipient's MTA in order to store messages in email boxes and retrieve them from there.*

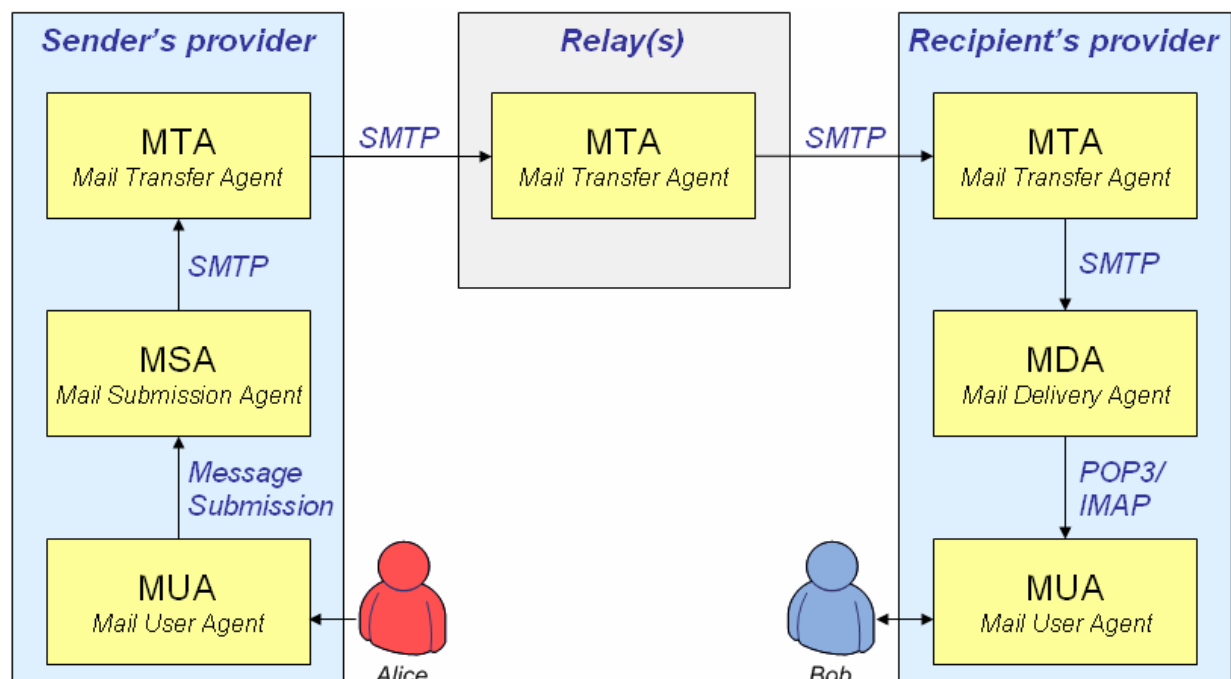


Figure 1: Composition of an email system

<sup>12</sup> See <http://www.sophos.com/pressoffice/news/articles/2007/04/dirtydozapr07.html>.

<sup>13</sup> See <http://www.spamhaus.org/statistics/countries.lasso> for the entire list.

The illustration describes the usual flow of an email from user Alice (left side) to user Bob (right side). Alice composes an email in her MUA and sends it ideally via the Message Submission protocol to her MSA, which forwards it to a trusted MTA via SMTP. Alice's provider's MTA will relay the mail to the MTA of Bob's provider. This can either be done via one or several email relays (as shown in the middle of the picture) or via sending the message directly to Bob's provider's MTA. In both cases SMTP is used as protocol for relaying the messages. Once the message has arrived at the recipient's MTA, it is passed to the MDA. The MDA is responsible for managing the user's mailboxes, e.g. for storing mails into them. In order to benefit from the asynchronous system of email communication Bob is able to access his mails later on by using his MUA and one of the standard mailbox access protocols POP3 or IMAP.

Since the email architecture is old and because this infrastructure is discussed often on the Internet, there exist some variations of this model. The most disturbing deviance, as defined (!) in RFC 2821 for SMTP is the leak of an explicit MSA in usual email systems. MUAs often skip the MSAs and send emails directly to an MTA without optimal authentication. Because there are only small differences between SMTP (protocol used with MTA) and Message Submission (protocol used with MSA), common MTAs adopted the incoming submission and are able to accept emails. However, this poor architecture weakens the reliability of emails, as described in chapter 3.1.

### 3 Current trends of spam

This chapter gives an overview of current spam trends as of the first half of 2007. It appeals to invest more time on defending spam and gives ideas about the future of some current spam methods.

#### 3.1 Spam via botnets

A botnet is a collection – more detailed a network – of autonomous computers, which can be controlled remotely via a specific application. Botnets are built from bad guys by using viruses, exploits and other threatening possibilities. The latter is a common way to create huge botnets that can be used by spammers. Those rent a botnet in order to spam from the participants, i.e. the single bots, and can control them via simple commands<sup>14</sup>.

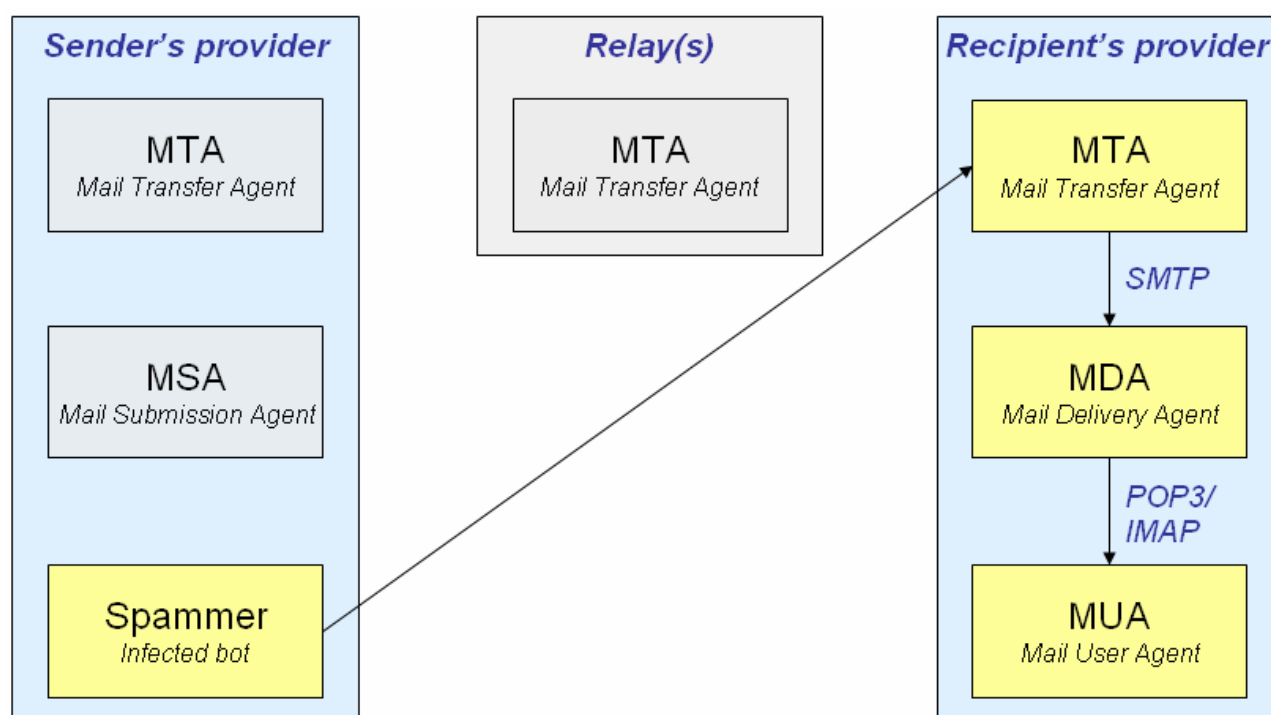


Figure 2: Spamming via botnets - circumvent MTAs

Frauds spam via botnets in order to disguise their identity and to avoid IP level filtering methods (as described in chapter 5.1). The method is very simple: Instead of sending legitimate mails via their provider's MSA or MTA, the infected clients (bots) deliver spam mails straightforward to the recipients MTA. Of the spammer's point of view this has mainly two advantages. In the first place they bypass connection level blocking of well-known spamming email relays by using "neutral" direct client connections to the recipients MTA. Firstly, it is more difficult for the MTA to identify the huge amount of bots that change mostly their IP addresses when reconnecting than the small amount of static spam sources. Secondly the spam trace begins at the infected bot and never includes a track with the IP address of the actual spammers, which is the (maybe temporary) controller of the botnet. If spammed via open relays or proxies, it would be difficult but possible to identify the spammer from an email's trace.

<sup>14</sup> See <http://honeynet.org/papers/bots/botnet-commands.html> for a large assortment.

About 80% of all spam is sent in this manner, i.e. via infected home computers that most often have a broadband connection to the Internet<sup>15</sup>. Missing awareness of home users leads to the fact that more than 25% of overall end-user PCs are infected with malware<sup>16</sup>.

Studies show contradictory behaviours of spammers that rent a botnet. Some observations show single bots sending a lot of spam (400 to 8,000 mails per hour) and justify that with the high costs of renting a botnet<sup>17</sup>. Other studies noticed an approach of sending less spam (100 mails per day) to avoid attracting attention to traffic monitors<sup>18</sup>. However, the huge quantity of infected computers allows spammers to use them when they feel like it and makes it possible to produce bulk email in short-time.

### **3.2 Localisation of spam**

In the early days of spam it was sent usually in English, in order to reach as many people as possible. Usually a non-English speaking user filtered out those emails, either manually when seeing the English language or by applying content filters which get used to score English emails as spam. Furthermore many of the recipients do not speak English well enough to understand the spammer's advertisement. Apparently spammers noticed this fact and try to localise their spam, i.e. translating the message into the language of the recipient.

Localisation can often be done very easily by either mapping the TLD of the emails to a language or by looking up statistics about usage of names in specific languages. For instance, an email address ending in ".pl" is likely to have a Polish speaking recipient, or an email address containing "Hans" is likely to have a German speaking recipient. Once this mapping between email addresses and languages has been done, spammers only need to translate the English mail into the most common languages (e.g. Spanish, French, German, Mandarin, Russian and Portuguese). The translation can be achieved by the help of automatic translators.

For the spammers localisation has two advantages. Firstly the recipient understands the localised spam message much better than an English mail. Secondly especially Bayesian filters miss spam emails in the recipient's language more often than emails in foreign languages (usually English), leading to a worse false positive rate. All in all localisation is a future trend with little effort but big benefit for spammers.

### **3.3 Image spam**

Since the majority of text spam is blocked by content filters, spammers look for new possibilities to cheat those. Thus, image spam is an approach, to disguise the text in an email through translating it into an image. Moreover, spammers commonly copy parts of well known websites into the body of the mail in order to fake its content and get good scores at Bayesian filters. The email's image(s) contains the actual advertisement.

But there are several more or less successful attempts to identify image spam. One possibility is using optical character recognition (OCR) at the content filter, which allows an automatic translation of the image back to text. After that, the text can be fed into a com-

---

<sup>15</sup> See [ENISA02], section 3.2, for details.

<sup>16</sup> See [http://www.darkreading.com/document.asp?doc\\_id=115563&print=true](http://www.darkreading.com/document.asp?doc_id=115563&print=true) for details.

<sup>17</sup> See [Comm01] or <http://www.mailchannels.com/research/spamonomics.html> for details.

<sup>18</sup> See <http://www-static.cc.gatech.edu/~feamster/publications/p396-ramachandran.pdf> for details.

mon content filter. Another effort is building checksums of the images or parts of it and comparing them with a spam checksum database.

Over the last months, spammers improved image spam severely. They manage to randomise the image for every single mail by using different possibilities<sup>19</sup>. To name but a few:

- Random pixels in the background
- Changes of colours (border, font, background)
- Multiple images appear as one image but impede OCR
- Non-static media (animated GIF, movies, ...)
- Image interference through “snowflake” patterns

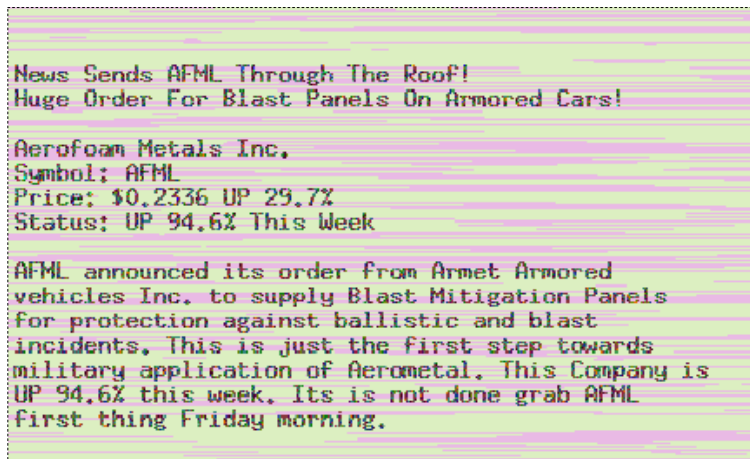


Figure 3: Example of concealed image spam

For this reason, anti-spam software mostly uses fuzzy OCR or fuzzy checksums to ignore these small modifications. Apart from that, spammers check images/emails against existing anti-spam tools, to test whether their blur is successful. If OCR was still possible, it would need much performance and costs a lot of resources<sup>20</sup>.

Finally one has to compare human readability with the potentialities of OCR. On the one hand, disguising an image too much will indeed exclude some readers from legibility. On the other hand, it ensures no OCR program can transform the image back to text. An interesting comparison can be made in the opposite direction: Many attackers try to pass CAPTCHA mechanisms<sup>21</sup> to register automatically at several sites. CAPTCHAs have the same intention as image spam.

As recently seen, spammers even use other formats than images to disguise their spam content. To an email attached PDF files or MS Excel tables are not readable (yet), and other formats like audio or even video files will help the spammers to complicate the detection of unsolicited emails.

### 3.4 Domain tasting

Spammers need a way how to sell their products, which is done on dedicated websites (except stock spam<sup>22</sup>). They take advantage of a five-day grace period to order domain names and to drop them at the end of the period free of charge. This method can be used to reduce or even avoid costs for a spamming campaign.

<sup>19</sup> See [Comm01] for a wide-ranging list of image spam permutations.

<sup>20</sup> Detailed discussion can be found on <http://lwn.net/Articles/196704/>.

<sup>21</sup> See glossary or <http://www.captcha.net/> for more information on CAPTCHA mechanisms.

<sup>22</sup> Stock spam is sent without any links and contains just the advertisement for a stock.

The period was originally introduced to rectify legitimate mistakes (e.g. mistyping the domain name), but is now widely misused. The responsible institution, ICANN, is still studying the problem and has not yet drawn conclusions from the fact, that in the year 2006 only 2% of all registrations were serious and long-dated (about 50% in 2004)<sup>23</sup>.

On top of it all spammers normally practice so called domain kiting. Domain kiting is a term for re-registration of the same domain name by the same wirepuller periodically. This can be done by a re-registration every fourth or fifth day, while each time making use of the grace period without ever paying for it.

### 3.5 Stock spam

Progressively spammers try to use the “stock dump” effect. The so called “pump and dump” is a kind of financial fraud that involves artificially inflating the price of a stock. Spammers try to push small traded stocks (known as “penny stocks”) after buying a huge amount shares. When sending large amounts of spam, the price of the shares increases, because some recipients buy shares. The spammer sells his shares at a good price and this is likely to fall again sharply.

From the spammers point of view this kind of spam is very efficient and safe. Spam can be sent with complete anonymity, because no link to any website is necessary. The spammer just has to state the stock and make an attractive offer for the reader. Those sometimes buy shares of this stock.

Three groups of individuals are involved in this business model:

- Spammers, who trade the stock to capitalise profits from their campaign
- Naive recipients, who believe in the pretended investment advises
- “Smart” recipients, who try to participate in price hikes triggered by spammers (though this is not recommended<sup>24</sup>)

As an example a common and freely chosen stock spam email has been evaluated. It was sent on Sunday, 22<sup>nd</sup> April 2007, to a German mailbox and advertised for the penny stock Country Line Energy Corp. with a starting price of 30 Cent per share. The market on opening rocketed from 30 Cent to a maximum of 37 Cent per share, which made it very attractive for the spammer to sell his shares. Interpreting the graph, the spammer actually sold his shares directly after the sudden increase. Over a long period, the stock price fell down to less than 1 cent per share. Similar documented observations can be found on the Internet<sup>25</sup>.



Figure 4: Share price graph of a stock, which was advertised via email spam

<sup>23</sup> [http://www.washingtonpost.com/wp-dyn/content/article/2007/02/18/AR2007021800599\\_pf.html](http://www.washingtonpost.com/wp-dyn/content/article/2007/02/18/AR2007021800599_pf.html)

<sup>24</sup> Spamnation set up a FAQ section, which among others discusses the possibility to benefit from stock spam campaigns. See <http://www.spamnation.info/stocks/FAQ.html> for details.

<sup>25</sup> Very detailed study from Rainer Boehme and Thorsten Holz is available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=897431](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=897431); Another Blog monitored a stock



## 4 Categorising anti-spam methods

Better evaluation of anti-spam methods can be made by grouping them into classes. Each category acts on a different level during the flow of ham and spam emails. In order to describe these categories, a reference model based on the Simple Mail Transport Protocol (SMTP)<sup>26</sup> will be developed.

### 4.1 Basics of SMTP

SMTP is the most common protocol used for sending emails, i.e. for the transport of data between MTAs or by the user sending an email from his MUA to an MTA. After the first introduction of SMTP<sup>27</sup> in 1982, the protocol itself changed only once in 2001. Between this, in 1995, the SMTP extensions<sup>28</sup> have been introduced, which allow some useful features like authentication or command pipelining.

Describing the groups of anti-spam methods can be done much easier, when considering the flow of SMTP. The following illustration shows a typical SMTP flow between a sending MTA (left) and a receiving MTA (right).

Sending MTA	Receiving MTA	
	220 foo.com Ready	} Email envelope
HELO bar.com		
	250 foo.com Hello...	
MAIL FROM:<xx@bar.com>		
	250 OK	
RCPT TO:<yy@foo.com>		} Email DATA
	250 OK	
DATA		
	354 Start mail input	
From: <xx@bar.com> To: <yy@foo.com> Subject: my email Date: Fri, 13 May 2007 13:33:37 +01:00 This is a text! .		
	250 OK	
QUIT		
	221 Good bye.	

The SMTP dialog can be split into two parts. In the first part, up to the DATA transmission, the email envelope is built up. This envelope is used for routing the email to the correct mailbox. After the DATA statement the actual data (header as well as body) of the email is transmitted, which is never used for routing. Applying this information for anti-spam methods, a reference model can be built up.

price during a stock spam period <http://sunbeltblog.blogspot.com/2007/01/and-this-is-why-you-see-so-much-spam.html>

<sup>26</sup> SMTP as described in RFC 2821; see <http://tools.ietf.org/html/rfc2821>.

<sup>27</sup> The "old" SMTP as described in RFC 821; see <http://tools.ietf.org/html/rfc821>.

<sup>28</sup> ESMTP as described in RFC 1869; see <http://tools.ietf.org/html/rfc1869>.

## 4.2 Architecture of anti-spam methods

With the background of the email transport protocols the anti-spam methods can be categorised into four types:

- **Email envelope analysis**

Once an SMTP connection has been established, at least four different kinds of information arrive before the actual data part:

- The IP address is given by the TCP/IP dialog and is very hard to forge.
- The sender's domain after the HELO, which can easily be forged.
- The sender's email address, which can also be forged easily.
- The arbitrary recipients email address.

Blocking spam at this part of the SMTP dialog is the most efficient possibility. It allows cancelling SMTP connections before receiving the emails' data and hence avoiding the traffic as well as the consumption of resources for processing and storing emails. To sum up, this behaviour usually blocks SMTP connections before receiving the actual email data. Therefore envelope analysis methods should check the data in real time in order to allow blocking just in time.

- **Email DATA analysis**

SMTP connections that passed the envelope analysis may deliver the email within the DATA part. After storing this email the data analysis can check the emails for conspicuous spam attributes. Very important is the fact that any information given at this place is arbitrary, though the pertaining RFCs define the original sense of the most headers. However, most spammers do not care about these standards and try to disguise their identity by inserting bogus data into the email header.

At this part data analysis methods interfere. They try to identify spam by applying algorithms on the email's header as well as content. On the one hand this procedure is less efficient than envelop analysis, because it needs much more resources. Usually disproportional more CPU time is needed and therefore these methods mostly cannot be applied in real-time. On the other hand it is a very helpful addition for current anti-spam installations.

- **Structural adjustments**

Some people gave up looking for anti-spam solutions for the current email infrastructure. Arguing against existing methods with bad quality, they try to solve the problem spam by deploying a silver bullet for it. These solutions are often incompatible to the current structure and require large changes. On the other hand they try to give a solution, which is less complex than the appliance of a mixture of current anti-spam solutions.

- **Useful anti-spam tools**

Differentiating explicit and implicit anti-spam solutions, this section describes the implicit ones. Every explicit anti-spam solution, regardless whether email envelope or data analysis, needs a database with defined rules. Therefore implicit anti-spam solutions exist in order to procure data for explicit anti-spam solutions.

The following graph serves as reference model for the composition of anti-spam methods:

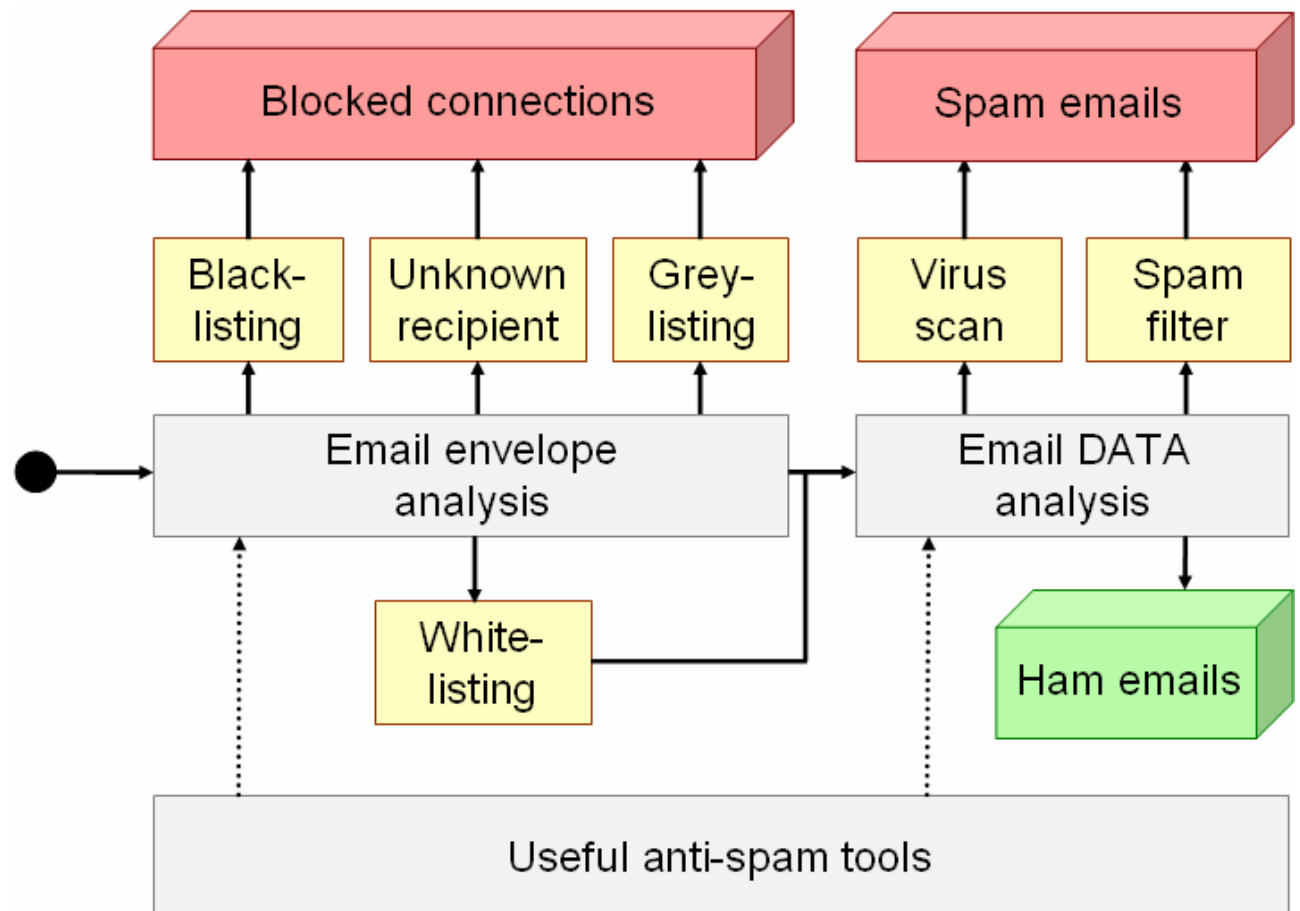


Figure 5: Reference model for anti-spam tools

*Useful anti-spam tools* are the fundament for both analysis methods and provide them with accurate and up-to-date data about spam trends. The flow of an email begins with the SMTP connection and the analysis of the email is split to an envelope and the DATA analysis.

*Envelope analysis* should be done at SMTP connection level, i.e. before accepting the DATA of an email. Thus envelope analysis leads to blocking whole connections, which would be probably used to spread spam. However, there are several discussions about the dilemma between benefit and risk of blocking whole SMTP connections, which will be further described in chapter 5.1.1.

Helpful methods to figure out bogus SMTP connections are (among others) black- and greylisting as well as shutting down connections for unknown recipients. Whitelisting is an efficient solution to avoid these methods being applied in order to guarantee the establishment of welcome communication channels. The specific methods will be described more detailed in chapter 5.1.

*DATA analysis* acts on per email basis and analyses each message. It categorises no longer SMTP connections, but rather individual emails into spam and ham. Several spam and virus filters can be used to achieve this. When an email has not been considered as spam, it is a ham email.

The fourth category does not fit into the context of this illustration. *Structural adjustments* change the kind of email traffic in such a manner, that – as far as applied correctly and comprehensively – usually no other anti-spam methods are needed. Therefore the process of integrating them into the current architecture of email turns out to be very difficult.

### **4.3 Different efficiencies of anti-spam methods**

It is important to differentiate the methods regarding their working points. Anti-spam mechanisms can operate at different levels. The earlier in the process of a mail delivery a mechanism is used, the more efficient the spam protection is. The vaguer a mechanism is the higher is the risk of categorising emails falsely into spam and ham. When comparing two different kinds of anti-spam methods, these differences figure out promptly.

Imagine an SMTP connection, which leads to 100 image spam emails. Any of these 100 emails passes among others at least through an SMTP dialog analysis, a virus scan and finally a content filter, which converts each image via OCR to apply a Bayesian filter on it. Apart from the wasted bandwidth for 100 image spam emails, many resources of the anti-spam software are used to check the content of the email. On the other hand, if an optimised network level anti-spam mechanism (e.g. DNSBL) was used, the SMTP connection would not be established. The last option should be preferred, since it saves bandwidth and other resources.

## 5 Overview of anti-spam methods

Many anti-spam methods are available for ESPs/ISPs. Tried and trusted are combinations of multiple implementations, which increase the efficiency a lot. This chapter gives an overview about the possible and most widespread anti-spam measures, which providers can take to reduce ingoing spam. Needless to say, each method has its pros and cons. These are explained for each type of anti-spam solution.

For each method concluding key facts are available. The symbols give an overview about the type of key fact. In detail these types are:



= Advantages



= Disadvantages



= Indications to danger



= Uncertainties

As far as dependencies on other technologies exist, these are also mentioned in this chapter. However, this chapter does not give a selection of best practice solutions. The extent of usage of the mentioned technologies is listed in chapter 6.2, a recommendation for appliance of technologies is given in chapter 8.

This chapter consists of four subchapters, each describing a category of anti-spam methods as discussed before:

- Email envelope analysis
- Email data analysis
- Structural adjustments
- Useful anti-spam tools

## 5.1 Email envelope analysis

Since SMTP is based on TCP/IP the sender's IP address within an SMTP dialog cannot be forged easily. Without using other extensions on SMTP, this IP address is the only quite reliable information. This part describes anti-spam methods allowing to block SMTP connections before receiving spam from them.

### 5.1.1 Blacklisting

In the context of anti-spam blacklisting describes the process of blocking upcoming SMTP connections from spammers, which are contained within a list of IP addresses (blacklist). The most common blacklists are DNSBL, which is an acronym for "Domain Name System Blocking List". The name results from the technique how these lists are used: When receiving an SMTP connection attempt the MTA usually requests via a special DNS record to a DNSBL whether the queried IP address is blocked or not. The DNSBL indicates this by returning an address (IP is listed) or a "NXDOMAIN" error code (IP is not listed). According to the answer the SMTP dialog might be aborted preventing a possible flow of spam.

Next to DNS based blacklists there are other proprietary types of lists on the Internet. Thus using them as a rule is more effort than integrating a standardised DNSBL. These proprietary types of blacklists are not as famous as DNSBLs. On the contrary, there are some providers taking the time to administrate an own additional blacklist.

Many DNSBLs are available on the market. Usually they are free, but some of them require a payment before using especially for high email volumes. It is possible to request multiple DNSBLs before blocking a connection attempt, which can lead either to better false positive rates (block only if at least X out of N blacklists suggest it) or to better false negative rates (block if one of N blacklists suggests it). Since there are several public DNSBLs, these can be categorised.

One approach is differentiating between the content of blacklists:

- Bogon/Hijacked net ranges
- Open form mailer
- Open relays
- Open proxies
- Dialup net ranges
- Other spam sources
- Mixed lists, i.e. aggregation of several above mentioned types

On the other hand, the type of inquiry for blacklists is interesting:

- Manually driven blacklists
- Automated input due to spamtraps or other methods
- Input via user submissions
- Combination of manually and automated input
- Aggregation of other blacklists

The risk of using a blacklist can and should be considered. Furthermore the up-to-dateness, the coverage of the whole IPv4 network address space, prices as well as add-

ing/removing policies have to be taken into account. Thus the available DNSBLs have to be chosen carefully. Chapters 7 and 8.1.2 include further discussions on blacklisting.



Blacklisting needs very few resources and protects against resource misuse, since email delivery is denied beforehand.



Blacklisting is independent from email's content, i.e. no liability to weakness of content filtering.



Adopting this method is very simple, even for multiple blacklists.



Risk assessment of blacklists is a very complex process and should be supported by experienced groups.



Blocking SMTP connections without looking into the emails might be dangerous, because no quarantining and in this way no recovery of false positives is feasible. Using blacklists for a scoring system instead of blocking of connections should be considered.



Condition of blocked IP might change more quickly than its reputation.

### 5.1.2 Greylisting

Greylisting as defined by Evan Harris<sup>29</sup> is based on the assumption that a legal email sender does more effort to send his email than a spammer. Therefore the server stores cookies for each connection attempt within a defined time span in the past. These cookies are triplets with the sender's IP address, email envelope sender address and email envelope receiver address. Generally – when greylisting – all connection attempts except those which already have a cookie are blocked. Moreover a cookie is installed at this step. This allows the sender to submit emails in a second try after a specific embargo time. Simply put, greylisting waits for a second attempt after a particular blocking time span before accepting an unknown sender's connection.

There are some improvements of this technique to increase the sender's efficiency. One of them requires within the triplet only an IP address from the same class C address block (which includes up to 256 hosts), because large senders often have a pool of machines sending mails. Another idea is to determine this pool of machines via Sender Policy Framework (SPF).



High benefit with very small effort, because many spammers often do not try a second time.



Spammers might adopt their methods if greylisting becomes more regular.

<sup>29</sup> See <http://www.greylisting.org/articles/whitepaper.shtml> for the full proposal.



Might interfere legitimate email traffic, if cookie does not exist before sending an email to a server using greylisting.

### 5.1.3 Whitelisting

Whitelisting is the opposite of blacklisting and prevents from anti-spam mechanisms being applied on the network level (i.e. grey- and blacklisting) for well known communication channels. A majority of (ham) emails is sent by well known sources, which do not need to be checked against grey- and blacklists.

Nevertheless, usually whitelisting a connection is by no means accepting every email from it. Other spam recognition methods as described later on should be applied on data delivered through the established SMTP connection. However, one could consider a solution allowing everything from this connection, which might lead to a very high risk of spam once a sender is whitelisted.



Whitelisting disburdens the resources which are needed for requesting blacklists and allows abstaining from interference of legitimate connections.



Whitelisting might correct false entries of remote administrated blacklists.



Once a whitelisted server begins sending spam, the trust to this server must be proofed if misuses occur.

### 5.1.4 Sender authentication

The email infrastructure grew up from a network of confident participants. Therefore from today's point of view, email has had a fundamental flaw from the beginning: a lack of authentication. In other words anyone on the Internet can, in theory, send emails to anyone else while claiming to be a third person. Thus sender authentication methods were born and have been widely discussed over the last years<sup>30</sup>. A very important goal of authentication methods is to keep existing email infrastructure and to not visibly affect current implementations. A change that would break existing mail clients or servers would be disastrous.

Two different kinds of authentication proposals can be found. Path-based algorithms watch where the mail was sent from. The best-known examples of path-based authentication are Sender Policy Framework (SPF) and Sender ID. On the other hand, signature-based algorithms determine whether the message is legitimate by using a cryptographic digital signature on the message. Following the main ideas of the two authentication methods are discussed, without getting to concrete and describing a specific authentication technology.

#### 5.1.4.1 Path-based sender authentication

Authentication methods like Sender ID and SPF can be used to test whether an email server is authorised to send on behalf of a given domain. The fundament for this is pub-

---

<sup>30</sup> The MTA Authorization Records in DNS (MARID) IETF working group tried up to 2004 to find a mutual consent about one sender authentication method. It ended without any results, because auf disagreements regarding the given sender authentication proposals.



lishing DNS records that list all authorised email servers for a domain. Now on the receivers site can be checked, if the domain of the given email address may be used by the sending server. This “given email address” differs between Sender ID and SPF: While SPF checks the envelope’s MAIL FROM, Sender ID tests the email headers.

### Disguising sender

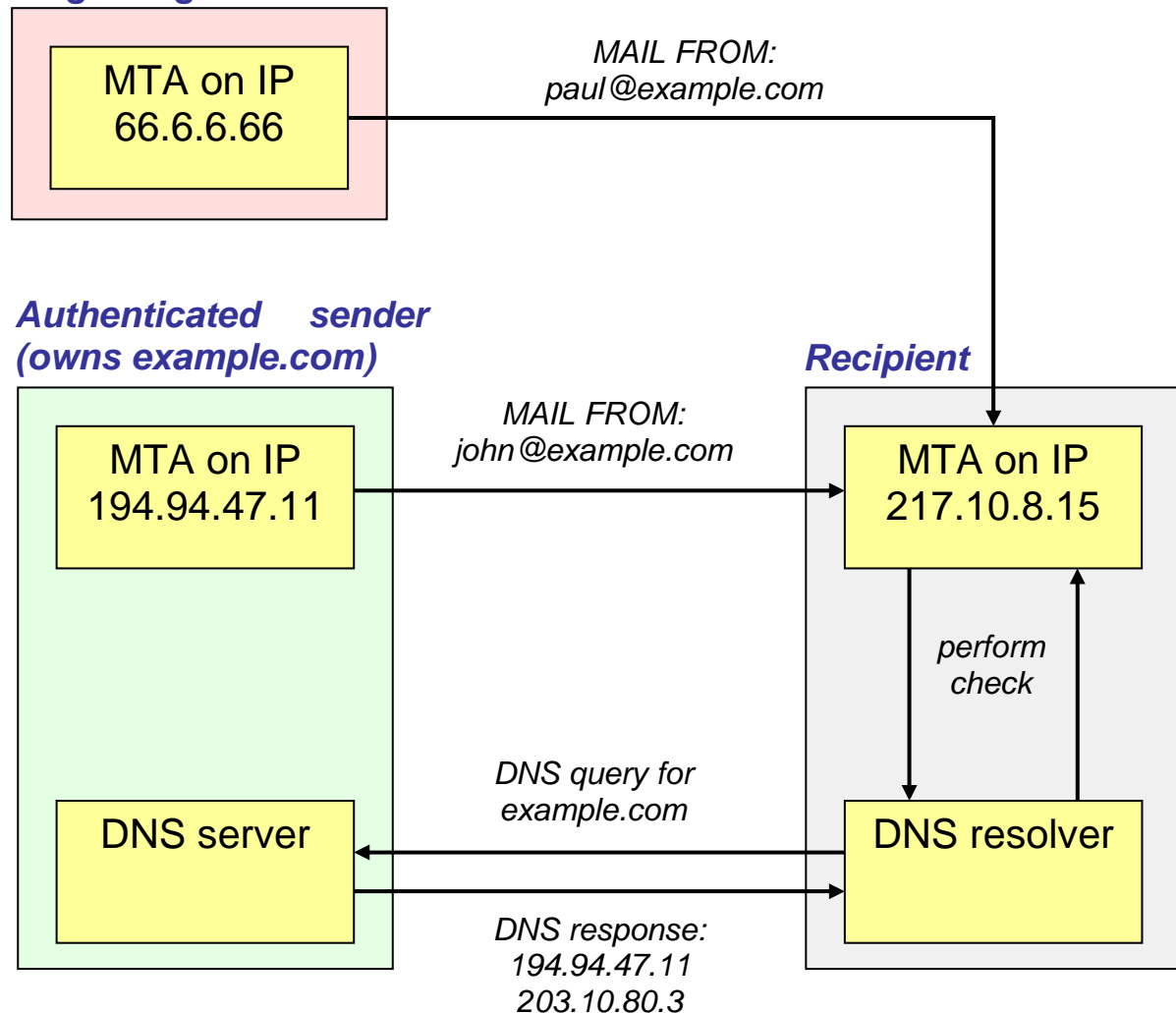







Figure 6: Data flow in path-based sender authentication

The figure explains two different situations. First considering the good case, i.e. an authenticated sender (green) owns the domain “mydomain.com” and sends an email from this domain. The recipient (grey) checks via a DNS resolver whether the sending IP is authenticated to send mails on behalf of “example.com”. The DNS server, which is responsible for this domain returns special<sup>31</sup> DNS records that express which IP addresses are authenticated to send mails from this domain. In the first case the sender uses one of the granted IP addresses (194.94.47.11) and is well authenticated. On the other hand, senders mailing from different IP addresses (red) are not authenticated.

<sup>31</sup> It depends on the actually method, which DNS record types are used.

-  Easily manageable for senders using existing technology, since only a DNS record has to be published.
-  Receivers have to adopt new software in order to check the DNS records.
-  Email forwarding services are generally prohibited when using this method. Two possibilities to manage these exist: Whitelisting of specific senders or rewriting of the senders email addresses (like Sender Rewriting Scheme<sup>32</sup> for SPF).
-  If two or more domains are running on the same IP address, an email sender from this IP can use every of those domains to send authenticated emails.
-  Prone to domain tasting, because spammers could use domains with legitimate DNS records.

#### 5.1.4.2 Signature-based sender authentication

Another idea authenticating a sender is using the (old) technology of a digital signature via asymmetric encryption. Existing implementations avoid the use of trusted third parties like certificate authorities. Similar to path-based sender authentication, but quite different from other authentication methods like PGP or S/MIME, signature-based sender authentication assures the use of the correct domain only, i.e. it is not possible to differentiate between several aliases of the same domain.

#### **Authenticated sender** (owns *example.com*)

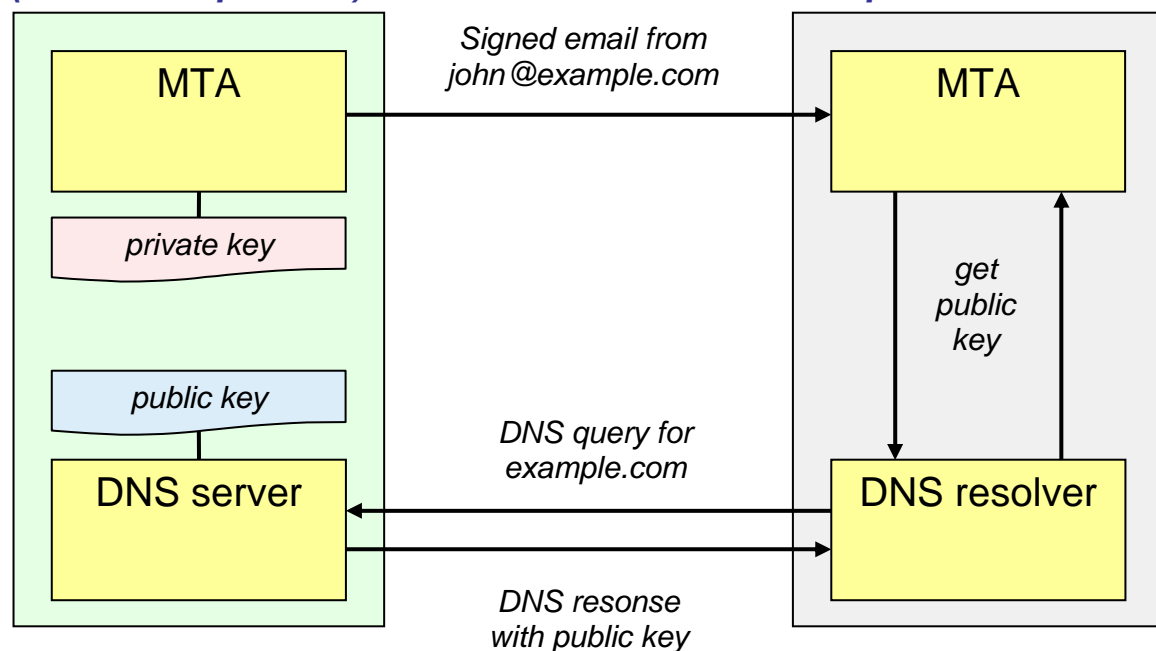


Figure 7: Data flow in signature-based sender authentication

<sup>32</sup> Sender Rewriting Scheme (SRS) is a mechanism for rewriting sender addresses when an email is forwarded in such a way that mail forwarding continues to work in an SPF compliant world. See <http://www.libsrs2.org/srs/srs.pdf> for more information.

The figure shows a typical situation for signature-based sender authentication. An authenticated sender (green) signs an email with a digital signature, using the private key available for his domain “example.com”. Usually the sending MTA signs the email, in individual cases this happens at the MUA. After sending the email, the recipient’s MTA (grey) retrieves the public key for this domain via DNS. It can use this key in order to check whether the signature of the sender is correct, just by decoding the messages signature and comparing it with the emails hash value.



Email forwarding is possible, without changing the sender’s address.



Differentiation between several domains hosted with a single IP address is possible.



The modification of emails is no longer possible, i.e. some software implementation must prevent this (e.g. mailing lists that put unsubscribe information to the end of every email).



Senders as well as receivers have to implement new technologies to sign emails and/or to check these signatures



Prone to domain tasting, because spammers could use domains with legitimate DNS records.







### 5.1.5 Sender address verification (SAV)

SAV, sometimes called *sender callouts* or *callout verification*, is a mechanism used to check whether an email address exists or not. Spammers often use mythical email addresses that usually do not exist. SAV helps only to verify if the sender’s given email address exists. On the other hand it does not help to verify if the sender is authorised to use this specific email address or domain. Since SAV helps only to block spam due to wrong email addresses and does not increase the reliability when receiving “verified” emails.

Technically the receiving MTA performs SAV with the given sender address during the SMTP dialog with the sending MTA. In order to do so the receiving MTA establishes an SMTP dialog to the MTA accepting emails for the domain stated in the sender address and tries to deliver a bounce message to this address. If the bounce message was accepted, the sender’s address (probably) exists and the receiving MTA should accept this email address. If the bounce message was rejected, i.e. a permanent error occurred (SMTP status codes 5xx), the email address is likely to be invalid.

The following example shows a sending MTA (left column) trying to submit emails to a receiving MTA (middle column). The receiving MTA rejects the email submission temporarily and sets up an SMTP dialog to the domain's MTA. Within this second dialog it checks via submitting a bounce message whether the email address exists. It quits the second dialog and is able to decide about the sender's address, which is in this case positive (the receiving MTA accepts it).

Sender MTA	Receiving MTA	Sender domain MTA
	220 foo.com Ready	} <b>Main dialog</b>
HELO bar.com	250 foo.com Hello...	
MAIL FROM:<xx@bar.com>	250 OK	
RCPT TO:<yy@foo.com>	<b>451 not yet verified</b>	
	HELO foo.com	
} <b>Sender address verification dialog</b>		220 bar.com Ready
		MAIL FROM:<>
		250 bar.com Hello...
		RCPT TO:<xx@bar.com>
		250 OK
	QUIT	250 OK
		250 OK
RCPT TO:<yy@foo.com>	<b>250 OK</b>	} <b>Continuation of main dialog</b>
	(...)	

-  Spam with wrongly spelled sender's email addresses can be filtered.
-  Will every MTA reject emails during the SMTP dialog if an email address does not exist or do some MTAs accept emails and create bounce messages?
-  Loops may occur when not using "<>" as sender address within SAV dialog and both MTAs performing SAV.
-  Spammers will probably learn from SAV and adopt their sender addresses to existing addresses, undermining the full benefit of SAV.
-  SAV might lead to DoS attacks against the owner to the used domain for spamming. High spam volumes with a specific sender domain will lead to many SAV checks performed by multiple parties, breaking down this specific MTA.
-  High resource consumption due to an additional heavy SMTP dialog.

## 5.2 Email data analysis

Once an SMTP connection is accepted and the email's DATA is delivered, this data can be analysed for spam-like patterns. Usually the following methods are applied after the SMTP dialog, i.e. after physically storing the email. Some implementations act during the data delivery, but checks during the runtime or the emails are likely to be too slow to be adopted just-in-time. Moreover quarantining/marketing mechanisms are desired, in order to allow recipients to check these presumable emails easily.

### 5.2.1 Heuristical methods

Often also known as rule-based content filtering, heuristical methods usually aim at finding specific words, regular expressions or misuse related styles in emails to classify them as spam or ham. Once conspicuous emails have been found, e.g. an outstanding expression of a spam email, it will be added as a new policy. A rulebook contains these checks and must be managed manually. A policy can either work on the email body or on the email header.

#### 5.2.1.1 Heuristical email header analysis

Spammers often perform very similarly when creating or modifying the header to spoof their identity or to optimise the email as ham. The origin of the email's native sender is ideally clearly readable from the "Received: from" lines within the email header. Originally, as described in RFC 2822 as "trace field"<sup>33</sup>, this data can be used to read the complete flow of the email message.

```
Received: from x.y.test
        by example.net
        via TCP
        with ESMTTP
        id ABC12345
        for <mary@example.net>; 21 Nov 1997 10:05:43 -0600
Received: from machine.example
        by x.y.test; 21 Nov 1997 10:01:22 -0600
```

This example, taken from RFC 2811 describing the Internet Message Format<sup>34</sup>, shows a typical scenario of an email trace. Each participating system during the transfer inserts its own information to the trace. A common chain could be:

$MUA_{\text{sender}} \Rightarrow MSA_{\text{sender}} \Rightarrow MTA_{\text{sender}} \Rightarrow MTA_{\text{relay1}} \Rightarrow MTA_{\text{relay2}} \Rightarrow MTA_{\text{receiver}} \Rightarrow MDA_{\text{receiver}}$

In order to disguise their identity, spammers insert several forged "Received: from" lines with bogus data that complicate identifying the origin of the email. The receiver can only trust the last "Received: from" before the mail has reached his trusted network. In the example above this would be the line added by  $MTA_{\text{relay2}}$ . Other "Received: from" information before this line could be forged by  $MTA_{\text{relay2}}$  and thus are not reliable.

Out of all reasons spammers often send emails with non-authentic "Received: from" information. They insert unknown hosts, do not care about date differences of several weeks

<sup>33</sup> See <http://tools.ietf.org/html/rfc2822#section-3.6.7> for more details.

<sup>34</sup> See <http://www.ietf.org/rfc/rfc2822.txt> for the complete RFC 2822.

between each line, or insert dates far in the future in order to be the most current email for a long time<sup>35</sup>. Heuristical methods can help to detect this misbehaviour.

### 5.2.1.2 Heuristical email content analysis

Because spammers tend to disguise the real word into multiple patterns, usually regular expressions are used to detect them. The following example of the most common word contained in spam email “Viagra” shows some possibilities to modify the word:

Viagra	=>	Vi@gra	Viraga
		V1agra	V_i_a_g_r_a
		Vi4gra	Viiagra
		Via_gra	Vi agra
		Wiagra	Vi<>gra

All forms make it more difficult to read the word, but nevertheless almost every human should be able to read it. It is much work finding a proper regular expression finding most of all possible cases.

In addition to this the list of bad words and/or regular expressions might get very long. This causes several checks for each mail and leads to a leak of performance. It is not recommended to quit the check after the first hit and classify the email as spam, because ham emails could contain some words or wrong headers, too. Hence first a combination of some positive checks should lead to a final classification.



The method does not need a training phase.



Heuristical analyses are very efficient with a well-managed policy database.



A leak of performance might occur with at high mail volume or huge policy databases.



Managing the policy list is very time-consuming.

## 5.2.2 Statistical methods







A statistical filter automatically splits emails into several tokens (e.g. words) and looks these tokens up in a database. The database contains common tokens with a classification whether or not it is a common token in spam emails. This requires a training phase of statistical methods, where lots of messages must be classified as spam/ham in order to build up the database.

Usually statistical methods are applied by the end user, i.e. within the MUA. In this document the end user’s possibilities to combat spam will not be discussed. But actually some approaches at the email server’s side exist, which will be debated.

<sup>35</sup> More detailed information and other mistakes made by spammers are described on <http://www.stopspam.org/email/headers.html>

Most of the statistic methods are based on Bayesian filtering<sup>36</sup>, which became popular in 2002. Implementing this at the provider enables to have a huge and best accurate Bayesian database. On the other hand a user cannot decide its own view of spam, i.e. users that tend to receive spam likely mails might have high false positive rates (e.g. researchers in the field of Viagra).

In addition to this, a method called “Bayesian poisoning” tries to trick the Bayesian filters by inserting ham likely word salad into the email. There are some discussions about this, which will be considered more detailed in chapter 8.2.1.

-  Automated method, i.e. after training phase usually no manually work needed.
-  Low false positive rate for text spam.
-  Training phase required, but at the level of providers almost negligible.
-  Bayesian poisoning might mislead the Bayesian filters.
-  Image spam has to be translated into text before analysis is possible.
-  Leak of performance at high mail volume.

### 5.2.3 Blacklisting of URIs (URIDNSBL)

Usually spam emails contain one or more links to websites, on which the advertised products are published<sup>37</sup>. This URI is the single part of the mail body that cannot be obfuscated. Other parts, like the text, HTML-Code, images etc. can be changed random in each mail to cheat statistical methods or heuristics. Therefore it is a good idea, to analyse the static domain of the URI.

URIDNSBL is an acronym for “Uniform Resource Locator Domain Name System Blacklist” and is similar to a simple DNSBL. Instead of looking up an IP address from where the spam is sent, the URIDNSBLs are used to look up tuples composed of domain and top level domain (e.g. “google.com”) as well as IPs, which are used by spammers within the mail to advertise products/services. The queried databases contain tuples used by spammers and thus are a memory of often advertised URIs.

Several free and non-free URIDNSBLs are on the market<sup>38</sup>. As a DNSBL, a URIDNSBL has to be chosen carefully before using it.

<sup>36</sup> See <http://www.paulgraham.com/spam.html>

<sup>37</sup> An exception for this rule is stock spam, see chapter 3.5.

<sup>38</sup> A good overview of URIDNSBL can be found on <http://www.dnsstuff.com/>.



Very efficient for spam mails with familiar domains.



Public domains (e.g. shortlink services) can be misused by spammers to avoid getting blacklisted with their own domain. Recursive queries could be a work-around, but would take some time.



No benefit for fighting spam without URIs (e.g. stock spam)

#### 5.2.4 Checksum comparison

The same spam email occurs in multiple different mailboxes, but a single recipient cannot decide whether that this email is being received by many other persons. The main idea of checksum comparison is to share unique fingerprints about received emails. This allows each member of this community to know if and how often a certain message was received by other users.

The unique fingerprints are not the email itself (because of data privacy and the high amount of data) but are generated by a checksum algorithm. Mostly these algorithms are fuzzy checksums (or “local sensitive hash functions”<sup>39</sup>) to avoid a modification of the hash value if only little modifications on the text have been made. This allows looking up an email even if the spammer modifies the content and the email in fact differs from other similar ones.

To avoid spammers checking their mails against checksums, the most checksum functions are proprietary, i.e. almost every checksum database uses its own function and is therefore quite incompatible with other databases. On the other hand some research has been made to find a generally usable function<sup>40</sup>, which enables a centralised database without caring about different algorithms.

As bulk emails are sent to many recipients, there is a high risk that they will be found in checksum databases, although the messages are wanted. Therefore a validation has to be made, either combining with other anti-spam methods or the end-user’s help. The help of end-users is very welcome on the one hand. On the other hand it is dangerous as well, since users tend to have bad false positive rates<sup>41</sup>.



Automated database filling possible via spamtraps possible.



Users can help to apply this anti-spam method, but classifying legal bulk email as spam might lead to high false positive rates

<sup>39</sup> For more details see a paper of E. Damiani et al., which describes the terms and work of local sensitive hash functions: <http://seclab.dti.unimi.it/Papers/pdcs04.pdf>

<sup>40</sup> The best known function is called nilsimsa. For more information see <http://ixazon.dynip.com/~cmeclax/nilsimsa.html>

<sup>41</sup> For instance: Some users incline to classify legal bulk email like newsletters as spam, if they don’t want to read it (anymore).





Algorithms are either proprietary (not comparable with other databases) or public (spammers can test them before modifying the mails).

### 5.3 Structural adjustments

The spam problem can be considered as weakness of the email architecture. Many approaches mentioned before try to avoid email abuse without changing the architecture. On the other hand many possibilities exist to win the battle against spam by adjusting the current email system, usually leading to big technical modifications or reforms of the communication procedures.

#### 5.3.1 Splitting message submission from message relay

SMTP is widely used as a message transfer as well as submission protocol. To distinguish these two terms please see the definitions of MSA and MTA in chapter 2.2.4. Put simply, message *submission* is the transfer from the MUA to a mail server system (client to server), message *transfer* the relay between MTAs (server to server). The IETF introduced a protocol for the message submission process<sup>42</sup> in December 1998. In April 2006 this has been improved and was replaced by the protocol Message Submission for Mail<sup>43</sup> as described in RFC 4409.

Message Submission for Mail splits up message submission from message transfer (terms as described above). The protocol is based on ESMTP and describes additional restrictions or allowances for this. Most important are two differences between message submission via SMTP or via Message Submission for Mail. The latter uses TCP port 587 instead of TCP port 25 to submit emails, which allows distinguishing between common SMTP relay functions and submitting mails. Moreover, most important, the new protocol requires SMTP authentication, what is not a standard within standard SMTP on port 25.

Authentication prevents users to send anonymous emails to other servers. Therefore it is possible for ISPs to forbid access to email servers without proper authentication. Since usually the provider's subscribers have no need to submit unauthenticated emails directly to foreign (i.e. outside the provider's network located) mail servers, this access restriction can be simply done by blocking access from all dialup hosts on their network to port 25. Doing so solves the problem of bots in the provider's network and therefore slashes the (outgoing) spam rate of the ISP. A bot has no longer access to foreign mail servers without authentication as provided in RFC 4409.

However, this strict approach needs to handle some exceptions:

- For the case customers need access to submit emails directly to MSAs *outside* the provider's network, they can do so using port 587 resp. Message Submission for Mail.
- If customers have set up own mail servers within the providers network, these can only relay mails
  - a) Via whitelisting this connection for using port 25 *OR*
  - b) Via relaying all emails to the provider's MTA, which relays it to the recipient's MTA.

<sup>42</sup> See RFC 2476 for details: <http://www.ietf.org/rfc/rfc2476.txt>

<sup>43</sup> See RFC 4409 for details: <http://www.ietf.org/rfc/rfc4409.txt>

Unfortunately, the majority of home users have no broad knowledge about Message Submission for Mail. This is caused by the outdated standard configuration of nowadays MUAs, which use SMTP or ESMTP on port 25 for submitting mails instead of the up-to-date protocol. Blocking access to this port would evoke troubles to MUAs connecting to mail servers outside the provider's network without using port 587. A draft by C. Hutzler describing Best Current Practices (BCP) for Email Submission is available at the IETF<sup>44</sup> and recommends mainly two actions for the transient until the complete establishment of RFC 4409:

1) *In order to promote transition of initial message submission from port 25 to port 587, MSAs SHOULD listen on both ports. MSAs MUST require authentication on port 587 and SHOULD require authentication on port 25 [...]*

2) *As delivered from the factory, MUAs SHOULD attempt to find the best possible submission port from a list of alternatives. That list SHOULD include the SUBMISSION port 587 as well as port 25. The ordering of that list SHOULD try the SUBMISSION port 587 before trying port 25 [...]*



Applying this method will reduce the providers outgoing spam volume to a negligible level, since most spam is sent via botnets connecting directly to foreign MTAs on port 25.



No loss of functionality, neither at client's nor at server's site



This method will not noticeably reduce the ingoing spam rate of the provider. Rather it helps to get a better reputation by other providers, since they receive less spam. If every provider would apply this method, the problem with spam via botnets was solved.



In order to be still able to use botnets, the bots are required to provide authentication when submitting mails. If frauds improve the technology of the bots, these could relay mails with a stolen, but apparently proper authentication of the user. However, this would make it much easier to identify infected computers by scanning their mail traffic for abusive behaviour.



Some customers will need support to adopt the changes to their mail servers (listing on port 587) and MUAs (using preferably port 587).






### 5.3.2 Proof-of-work

The intention of proof-of-work methods is to prove that the sender did some work before sending an email. For instance this work could be bandwidth consumption or spending processing time. Since the sender has to do this work before, the receiver can guess his seriousness. The goal of proof-of-work is quenching illegitimate senders with a high work

<sup>44</sup> See the last version of this draft (dated 30.05.2007) at <http://www.ietf.org/internet-drafts/draft-hutzler-spamops-07.txt>




load. Proof-of-work should only be used in combination with a whitelist to save the work load for legitimate senders.

The most well-known implementation of proof-of-work methods is Hashcash<sup>45</sup>. It is based on calculating a specific hash value, which ensures the sender has spent some time before sending the mail. Hashcash can easily be integrated into the SMTP protocol without interfering applications that do not understand the procedure. Once a sender is validated via Hashcash he is put on a whitelist of legal senders and does not have to pass this procedure a second time.

-  Illegitimate senders with high volume are decelerated.
-  Incrementally deployment is possible, i.e. header can be ignored if not understood.
-  Do botnets really have leak of performance to do this additional work?
-  Different hardware and Moore's law<sup>46</sup> can threaten this method, because the runtime needed to proof the work differs between the individual machines.
-  Legitimate bulk email (newsletter, mailing lists) senders or too low-end hardware cannot either bring this work up or lack of using the proof-of-work mechanisms, i.e. every recipient has to whitelist the mailing list sender's address.

### 5.3.3 Challenge-response mechanisms




A challenge-response system expects a sender to react on a challenge before sending an email. This challenge itself is an email containing instructions how to respond in order to send the first mail successfully. Because of this complex procedure, challenge-response systems are often being combined with other anti-spam methods and challenge only if there is a suspicion for spam. For simplicity reasons it should be guaranteed answering an email does not lead to a challenge and is delivered immediately instead. CAPTCHA mechanisms within the challenge could make it much more difficult for spammers to send emails automatically.

-  Complex procedure combined with CAPTCHA makes it very efficient.
-  Acceptance by users?
-  Email backscatters can be used for DoS attacks.

---

<sup>45</sup> See <http://hashcash.org/>






<sup>46</sup> "The complexity for minimum component costs has increased at a rate of roughly a factor of two per year [...]", i.e. steadily increase of computers' performance.

-  Legitimate bulk email (newsletter, mailing lists) senders cannot stand this method and have to be whitelisted.
-  High traffic due to outgoing messages.
-  Long procedure for sending an email.

#### 5.3.4 Electronic postage

The cost of each email is almost entirely borne by the recipient of a message. Traffic-, storage- as well as backup costs boost the expenses for receiving and storing emails. On the other hand, sending bulk emails is very cheap and easy. If senders had to pay a small amount for every email, i.e. comparable to the current snail mail system, it would prevent them to send as many emails as they do currently.

All proposals of e-postage are more or less based on buying stamps before dispatching an email. This fact requires a micropayment system to handle the exchange of value for e-postage. After stamping, the email can be sent to the recipients. Usually the proposals are designed to refund the postage of legitimate emails, if the email is considered as ham. However, there are several arguments against electronic postage, both technical and social<sup>47</sup>.

-  Economics of spammers are broken.
-  Do users accept this new system, since they usually dislike micropayment systems<sup>48</sup>?
-  Danger of exploiting the micropayment system.
-  Users could decline the refunding, even if the email is ham.
-  The micropayment infrastructure is very difficult to manage, i.e. the financial, administrative and social costs of e-postage are completely unknown.

#### 5.3.5 Traffic shaping

The basic idea of traffic shaping is throttling the connection between an email sender and a mail server, in order to force impatient senders dropping their messages. Spammers tend to send emails very fast (as described in chapter 3.1). As a conclusion they quit connections and go to the next one, if they cannot deliver the mails in a short period of time. Traffic shaping is mostly realised by using an SMTP tarpit or the possible SMTP greeting delays.

<sup>47</sup> See <http://www.taugh.com/epostage.pdf> for a more complete list.

<sup>48</sup> See <http://www.dtc.umn.edu/~odlyzko/doc/case.against.micropayments.pdf>

Some deliberations think about interrupting the economics of spam in this way. Background is the fact that spammers most rent botnets for sending spam from them and of course pay for it. Once they have a botnet, they try to send as much spam as possible to bring a return for the payments. As this process would be globally slowed down, the spammers got problems to yield profits.



Global effect if everyone applies it.



Most spammers are impatient and abort slow connections.



Negative consequences for legal senders (mails are delivered slowly or legal sending mail server does not accept this slow speed).

## 5.4 Useful anti-spam tools

After describing the most important anti-spam methods, appropriate tools which provide accurate data for fighting spam have to be discussed. These tools increase the efficiency of the methods mentioned before. They decrease the false positive and false negative rates of these measures.

### 5.4.1 Spamtraps

Spamtraps are mailboxes dedicated for collecting spam. Thus they are no anti-spam method by itself, but more a very efficient tool supporting other anti-spam methods. Dealing as a special kind of honeypots they attract spam. As a positive effect receiving dedicated spam to an account can help to automatically scan through the conceived data. Many databases can be fed with attributes of the received spam.

- Saving the senders IP address in a blacklist or the like helps to detect more spamming sources on the Internet.
- Building hashes of the email supports building up databases for checksum comparisons to other emails.
- The text of the email can be used to tokenise it and use this information for statistical filtering (e.g. Bayesian filters)

Automated use of this data is only accurate, if nothing but spam hits the spamtraps. To make spamtraps useless, spammers try to disrupt this spam flow by sending ham emails from legitimate senders to a spamtrap. This causes the spamtrap to analyse ham as spam and feeding the database with wrong data. Once such a case occurred, further use of this undermined spamtrap cannot be recommended anymore.



If not exposed, data can be perfectly used for analysing spam.



Little effort necessary for running a spamtrap with huge benefit.



Vulnerabilities if spammers detect spamtraps and misuse them.

## 5.4.2 Reputation systems

Many anti-spam solutions rely on qualified base data to make decisions whether an email is likely to be spam or not. This information can either be shared around the Internet or it can be for a provider's private use only. The first option is of course much more efficient, since sharing data in a network provides a multiplier to the volume of data. However, keeping the quality of data reasonable requires some thoughts on the work and structure of anti-spam reputation systems.

Reputation systems in the field of anti-spam have many use cases. The most obvious and simple ideas about exchanging data are:

- **Dialup ranges of providers**

Many providers will not ever block outgoing mail on port 25 (as described in chapter 5.3.1). As recommended to providers in chapter 8.1.1, it is then very useful to share their dialup ranges with other providers in order to minimise spam via botnets.

The shared data can be checked by the recipients via looking up the data in their routing tables. If the received net range is routed to the sender's ASN, the information is verified and can be considered as correct. Thus exchanging dialup data is very secure and reliable.

- **Blacklisted IP addresses or addresses recently seen sending spam**

Due to frequency analysis or spamtraps providers figure out many IP addresses used by attackers in order to send spam. Probably these addresses will again send spam later on. Depending on the kind and the aggression of the spam received these addresses can be used to block or adversely affect future email communications with these partners.

Distributing such kind of data needs some considerations before. What is the likeliness an IP address is only sending spam and no ham? In which level do I trust information distributed by my partners? Who is reliable enough to deliver me this information I will use in sensitive applications like anti-spam? A sufficient reputation system should be able to handle these questions by configuring it accordingly.

- **Whitelisted IP addresses**

When using blacklisting, implementing a whitelisting mechanism is unavoidable. Thus combining the exchange of blacklisted IP addresses is similar to the exchange of whitelisted addresses. A provider should for political reasons consider whether it wants to share all its whitelist entries or exclude some from distribution.

- **Hash values of spam**

In order to create a comprehensive database for checksum comparisons to spam emails, the hash values estimated via spamtraps or other reliable sources are appropriate for an exchange. The actual data of the email is completely disguised, so this method is perfect for keeping the privacy.

- **Regular expressions for finding spam**

As described in chapter 5.2.1 regular expressions help to filter out spam. Once a regular expression is developed and tested, it can be spread via a reputation system. This minimises the work of email administrators, since for each spam issue only one partner of the system's network needs to spend time on creating an expression.

After showing some ideas of a reputation system's content it is important to choose an appropriate structure for reputation systems. Mainly there exist two approaches of reputation systems, each with its individual characteristics:

- **Centralised systems**

This approach saves all reputation data in a central database. Administrators or all network participants have the possibility to add, modify or delete content of the database. This structure is much easier to develop and to control than a distributed system. On the other hand these systems lack of robustness due to a single point of failure.

- **Peer-to-Peer networks**

In order to increase the availability of systems Peer-to-Peer networks can be used to exchange data via a reputation system. In a Peer-to-Peer network the peers are connected directly to each other without using a central server. These systems are more difficult to implement and lack of an absolute control. On the other hand they offer some features individualising a networks peer's connection. In other words controlling with whom or under which conditions a participant peers with other partners is much easier.



Sharing knowledge enhances the efficiency of anti-spam tools.



Only little effort when joining a network and contributing to a reputation system.



Can the performance of a reputation system keep up with mass data?



The quality of reputation systems has to be well-considered.



Developing and managing a reputation system might be a high effort.

### 5.4.3 Frequency analysis

Providers have the possibility to distinguish common traffic from abusive usage of the system. Constant observations of the network can lead to accurate data of misuses. This helps to increase the efficiency of legal as well as technical battles against spam. As good this sounds, as difficult it is to describe and evaluate the possibilities of frequency analyses.

Usually frequency analyses monitor the network or email traffic and watch for anomalies. This can happen in several ways, but each kind of analysis should be combined with

whitelisting in order to avoid analysing legitimate traffic. Some ideas of traffic analysis could be:

- How many emails does a specific connection partner, identified by its IP address, send in a specific time period?
- Does a connection show abnormal behaviour during the SMTP dialog? In detail:
  - Does it tolerate slow communication when using a tarpit?
  - How many recipients does the mail have?
  - Are some of those addresses wrong/undeliverable?
- Does a sender respect negative SMTP response codes finishing the dialog?

In order to keep clear of mixing frequency analysis with content analysis (i.e. email data analysis as described in chapter 5.2) frequency analysis as described here only uses the message envelope data.



Big potential of recognising spam at a quite different layer than other methods.



Helps to increase the perspective of anti-spam tools, since they usually consider only a single email.



Might be high effort to establish a reliable frequency analysis.



## 6 ENISA survey on anti-spam

ENISA<sup>49</sup>, the European Network and Information Security Agency, is an agency of the European Union working for EU institutions and EU member states. ENISA started its operations in September 2005 and is located nearby Heraklion (Crete) in Greece.

The agency's mission is essential to achieve a high and effective level of network and information security within the European Union. Together with the EU institutions and the EU member states, ENISA seeks to develop a culture of network and information security for the benefit of citizens, consumers as well as business and public sector organisations in the European Union.

As part of its Work Programme 2007<sup>50</sup> ENISA had to survey electronic communication measures after doing so one year before in 2006. In detail ENISA should report the analysis of technical and organisational measures that providers take to secure their networks and services from spam and other security threats. The main goal was to develop a sense of best practice among providers regarding such measures.

While working on this document, the ENISA study for the work programme 2007 has been deployed and its results have been evaluated. The full report created by ENISA was as of August 2007 still in work and will be published on the ENISA website approximately in October 2007.

### 6.1 Description of the survey

One goal of the survey was to give a sense of best common practices and figuring out potential adjustments of the EU or the member states' legislation. The questionnaire concentrated on technical and organisational measures applied by providers in order to ensure their networks security. Provider in this case is a generic term for ISPs, ESPs and telecommunication companies, which were contacted during the study. As described in the ENISA Work Programme 2007, the survey concentrates on spam and other security threats. Therefore this questionnaire consists mainly of two parts: Firstly questions about general network information security and secondly specialised issues on anti-spam measures. The full questionnaire is available in annex C.

In early June 2007 the questionnaire was spread to 920 direct provider contacts. Additionally it was distributed through large anti-spam coalitions like MAAWG<sup>51</sup>, euroISPA<sup>52</sup>, ETNO<sup>53</sup> and ECO<sup>54</sup>. Potentially participants had the chance to provide their responses either via an MS Word document or via an online web survey until 30<sup>th</sup> June 2007. Since ENISA's scope is the European Union, primarily providers of the EU member states<sup>55</sup> were contacted. However, some contacts outside the EU were used in order to compare both groups and their results.

---

<sup>49</sup> For very detailed information see the ENISA website at <http://www.enisa.europa.eu>

<sup>50</sup> See [http://www.enisa.europa.eu/doc/pdf/management\\_board/decisions/enisa\\_wp\\_2007.pdf](http://www.enisa.europa.eu/doc/pdf/management_board/decisions/enisa_wp_2007.pdf).

<sup>51</sup> Messaging Anti-Abuse Working Group, see <http://www.maawg.org>

<sup>52</sup> European ISPs Association, see <http://www.euroispa.org/>

<sup>53</sup> European Telecommunications Network Operators' Association, see <http://www.etno.be/>

<sup>54</sup> See <http://www.eco.de> for more information.

<sup>55</sup> For a complete list of these see [http://www.enisa.europa.eu/pages/country\\_pages.htm](http://www.enisa.europa.eu/pages/country_pages.htm)

## 6.2 Discussions of the results

After closing the deadline 30 different providers from 19 different countries responded to the survey. The survey represents a large part of the European Union with answers from 16 EU member states. Furthermore some big providers, in detail two of the top-three respectively three of top-ten biggest European broadband service providers<sup>56</sup> delivered high quality data. This ensures representative data for other European service providers.

### 6.2.1 Spam is a critical security threat

In order to get impressions of the participants about most concerning threats for their organisation the questionnaire<sup>57</sup> included in question one a part asking for ordering eight possible threats from most to least concerning. Spam was just after viruses the 2<sup>nd</sup> highest internet security threat when calculating the average of all choices. Less than a fifth of all participants considered spam as no problem and chose it as one of the last three concerning threats.

The third most feared threat is Denial of Service (DoS). This might lead to the conclusion, that botnets are in fact one of the biggest dangers on the Internet. Once infected by a virus, a computer within a botnet is usually used for spamming or DoS, a combination of the voted top-3 threats.

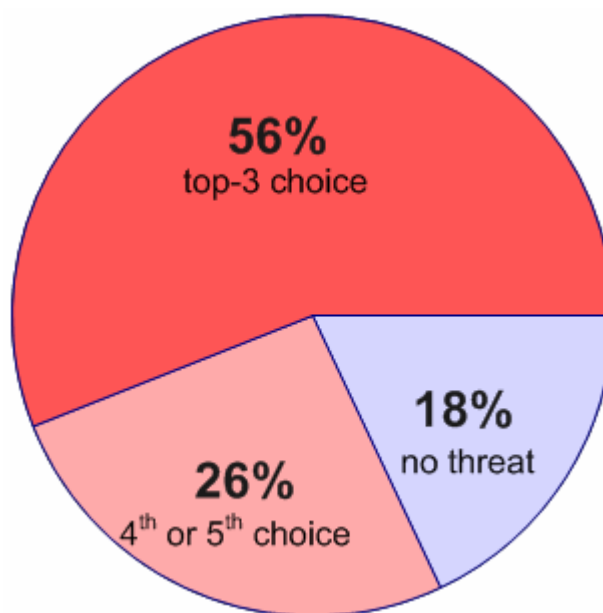


Figure 8: Spam as a security threat

### 6.2.2 Analysing spam

Generally a provider can take proactive and reactive measures to analyse where spam comes from. Two out of three participants react on requests from other providers who received spam from their network, which seems to be rather a bad than a good average. The same amount of providers also analyses consecutive complaints from their customers. Concluding from these two figures a third of all providers does not analyse where spam comes from when receiving an external complaint.

One third of the participants studies spam also when an automatically monitored spam level reaches a certain threshold. This kind of proactive monitoring is very efficient and avoids long latencies of complaints. Frightening is that 15% of all providers do not analyse where illegitimate email comes from at all and give spammers in this way the possibility to spam via their networks without taking action against it.

<sup>56</sup> A complete list of the top-75 biggest European Broadband Service Providers is available at <http://www.strategyanalytics.net/default.aspx?mod=ReportAbstractViewer&a0=3482> (fee required).

<sup>57</sup> See annex C for the complete questionnaire.

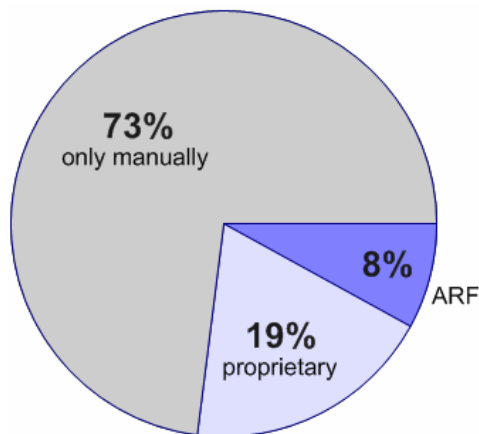


Figure 9: Using of automated processing of abuse reports

Abuse reports are usually sent to the email address `abuse@domain` and processed by the providers. With 73% the majority of all participants process the abuse reports only manually. Only two providers work with the abuse feedback reporting format (ARF), which is as of July 2007 an IETF draft in order to automate abuse reports<sup>58</sup>. However, especially bigger providers try to automate their process either via ARF or other, proprietary developments.

Especially huge vendors are able to provide feedback loops to bulk senders. A feedback loop relays information about abuse complaints regarding a specific email to the origin of this email in order to allow him to improve the bulk sending quality. Three big providers stated to provide feedback loops.

### 6.2.3 Most used spam-filtering measures

Reviewing anti-spam methods in use<sup>59</sup>, providers usually apply IP blacklisting (82%) and content filters (75%). Alarming is the fact that a high amount of more than half of all blacklist users do not utilise a whitelist, which is recommended as described later in chapter 8.1.2.2. Very rarely a provider uses an outsourced system in order to filter spam, which is surprising considering the high efforts for small providers. Next to the two main methods providers build on greylisting and filtering due to failed sender authentication (both 50%).

More advanced techniques like checksum analysis and reputation systems are employed only in less than a third of all providers. Blacklisting based on URIs as a special kind of content filtering is used by approximately two out of five providers. Also interesting is the amount of 43% of providers slowing down a sender's connection as a kind of traffic shaping. Only about a fifth of all providers use frequency analysis as a measure against spam. Paradoxically rather smaller than bigger providers avail themselves of this method, gaining the more power the bigger the network is.

The average number of methods a provider uses is 4.7, showing that there is no silver bullet for solving the spam problem. Only a good combination of anti-spam methods leads to a sufficient result in the combat against illegitimate email.

### 6.2.4 Efficiency of anti-spam methods

It turns out that blacklisting is not only the most widely used. It is moreover the most efficient anti-spam method. Aggregating the data mentioned by the biggest five participants of the survey, blacklisting had an efficiency of 70% averaged. When damping aberrations by building a median, blacklisting even filtered out 80% of all ingoing SMTP connections. Other SMTP-level methods like greylisting, whitelisting or filtering of unknown recipients were considered less efficient, but collaborated together with blacklisting in a good way. All network level methods summed up lead to decisions in 90% of all SMTP connections.

<sup>58</sup> For more information on ARF see <http://mipassoc.org/arf/>.

<sup>59</sup> See chapter 5 for details.

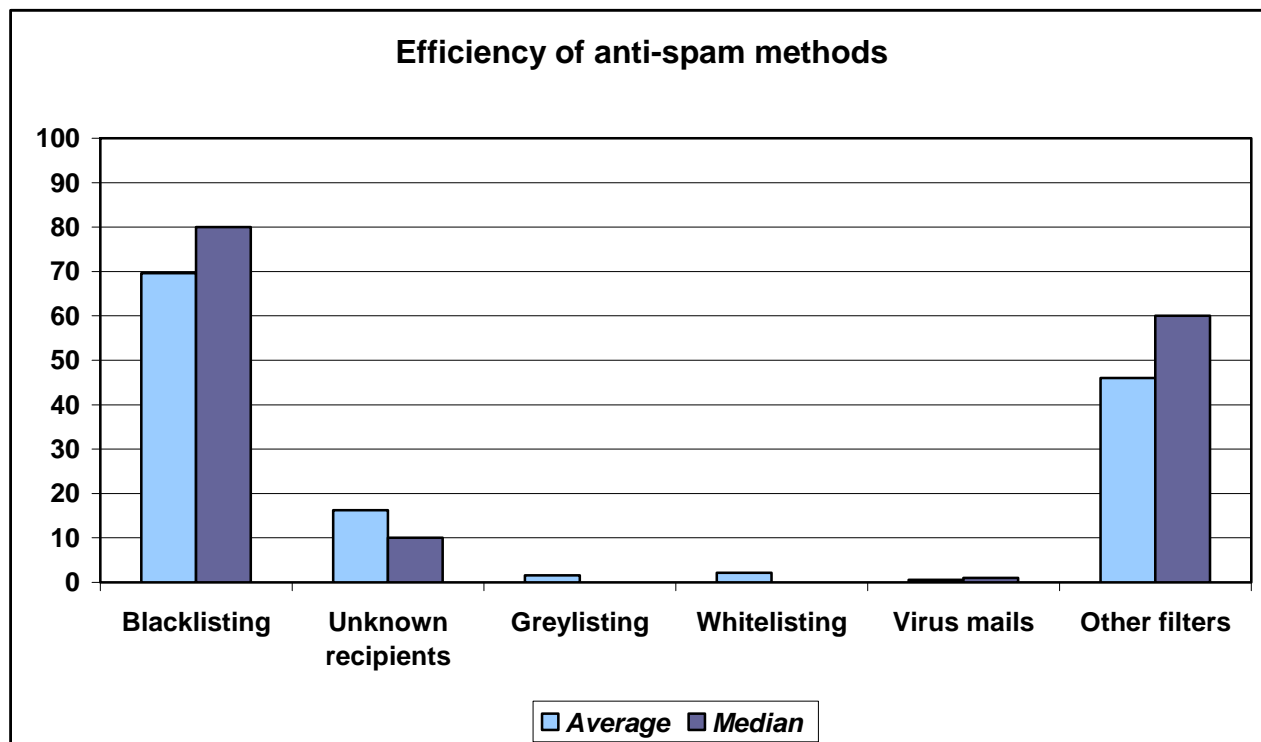


Figure 10: Efficiency of anti-spam methods

Greylisting was only used by one of the five big providers with a low efficiency of 8%. The average of virus mails was very low with 0.6%, confirming the trend analysis of other studies predicting even lower rates in the next years. Two of the five providers used whitelisting and privileged 5% respectively 6% of all connections.

On average 87.2% of all connections get blocked due to network level mechanisms (black- and greylisting as well as unknown recipients). Therefore only 12.8% connections lead to the entire amount of emails. All top-5 providers used other filters, which include methods described in chapter 5.2. These methods filtered medium 46% of all mails, which passed through the SMTP based filters before. Combined with virus filters only 53.4% of all accepted emails were not recognised as spam.

Assuming that every SMTP dialog leads to at least one email, the following formula calculates the efficiency of all anti-spam methods combined:

$$\text{Emails passing} \leq \% \text{ of accepted SMTP dialogs} \times \% \text{ of non-filtered email}$$

In this study only less than 12.8% x 53.4% = 6.9% of all planned messages managed to pass through the anti-spam tools. This is a satisfactory number, but shows on the other hand that at least 93.9% of emails were not accepted. The scale of the problem of spam is dramatically high, since roughly only one out of twenty emails seems to be ham.

### 6.2.5 Review of sender authentication

As mentioned earlier about a half of the polled providers use sender authentication as a means for spam filtering. Organisations provide authentication information, although they might not use such data for spam filtering. SMTP AUTH is a de facto standard for

authentication and applied by more than 80% of all participants. POP3 before SMTP complements this security mechanism with 19% usage.

Results concerning authentication standards are not that clearly evident. The discussion within IETF regarding authentication mechanisms led to confusions and diverging perceptions. This becomes clearly visible when reviewing the results of the survey, where none of the authentication mechanisms (as described in chapter 5.1.4) were applied in more than half of all cases. Methods that are implemented easily like SPF (50% usage) and Reverse MX (28% usage) are used rather than more complex solutions as DKIM (6% usage). On the other hand, two providers (one bigger) mentioned they plan to implement DKIM, justified by the latest standardisation of DKIM in May 2007.

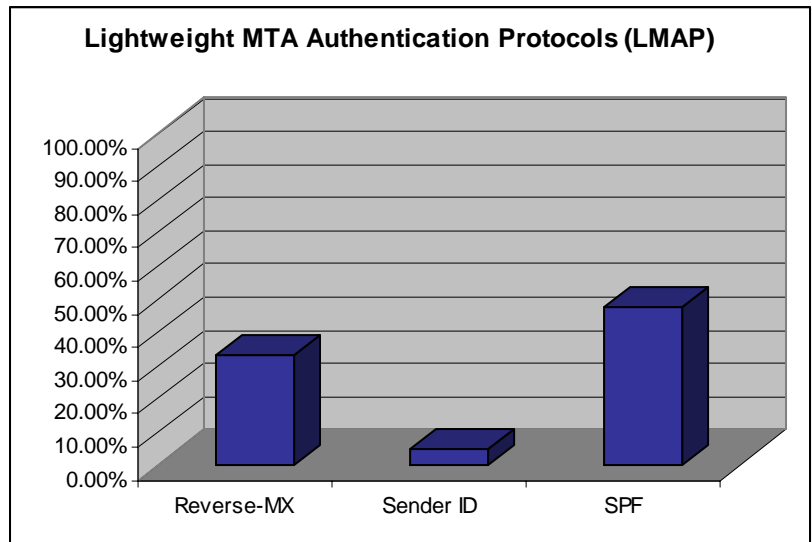


Figure 11: Comparison of LMAPs after MARID's proposals

An obvious trend is the decreasing usage of Microsoft's Sender ID which was used by only one provider. Sender ID seems to have lost the race versus SPF, which are competitive mechanisms of path-based sender authentication. Apart from that Reverse MX is still a serious contestant of SPF, pointing out the need of a single standard in this field.

### 6.2.6 Protection against outgoing spam

Providers do much work in order to protect their networks against incoming spam. On the other hand, outgoing spam does not concern them directly – as consequence providers could do only the bare necessities. But this might lead to bad reputation within the worldwide anti-spam scene with bad ramifications.

Most of the providers (88%) apply egress filters on their network to mitigate security threats like spam. Nearly a half of the participants block access to port 25 from all hosts on their network<sup>60</sup>, which is a very important step to fight botnets regarding spam. On the contrary only 24% provide Email Submission services on port 587. This would be an important step in order to force the customers to authenticate before sending mail. Half of the conducted providers limits high outbound email volumes, which is especially useful for free or anonymous mailing services. Two thirds of the providers put subscribers on a blacklist if they repeatedly send spam.

Concluding the providers actually apply measures to protect against outgoing spam. But especially the high efficient measure of managing port 25 should be used more widely in order to mitigate the high email volumes originated in botnets.

<sup>60</sup> Possible when applying Message Submission as described in chapter 5.3.1.

## 7 Empirical research on blacklisting

Blacklists are a very important tool in the battle against spam. Most people rely on hearsay or feelings when choosing blacklists. As always this might be an input for a part of the choice, but should not be the only one. This chapter introduces methods how to review the quality of blacklists and supports to decide on how or which blacklists should be used.

Moreover this chapter publishes research results on intersections between blacklists, on regional biases of those or on top-listed spamming providers represented by Autonomous Systems. It might help providers to see their current reputation in terms of facts, i.e. what ratio of their networks is listed by which blacklist.

### 7.1 Description of research methods

Researching the large number of available blacklists for email is almost impossible. Thus this research considers exclusively DNS and IP address based blacklists that offer a service to synchronise its data via common tools (e.g. `rsync`<sup>61</sup>, `wget`<sup>62</sup>). Moreover only the most well-known blacklists were reviewed, taking information from Spamlinks.net<sup>63</sup>, former experiences and various information exchange points like the Usenet into account. Unfortunately some blacklist operators did either provide a non-processible format (i.e. not convertible to the well-known `rblDNSD` format<sup>64</sup>) or the provider required a process paying money. In both cases the blacklists could not be integrated into this research.

Once the raw data of a blacklist was accessible it was imported to a relational database. By performing complex SQL queries on this database it was possible to build the statistics in a way with a very high performance. The results of these queries were taken from sample data on 12<sup>th</sup> July 17:00 p.m. UTC. They might be representative for a longer time, but for accuracy reasons ongoing research should provide this data on a daily basis.

### 7.2 Origin of data

Usually blacklists are a collection of different listing reasons. The following table outlines the used blacklists, the type of data included and a link to the policies.

Table 2: Explanation of data used for further research

Blacklist	Content	Policy
<b>all.dnsbl.sorbs.net</b>	Aggregated list of open proxies, open relays and other blocks. As of 11 <sup>th</sup> July SORBS did administrate neither the dynamic user host (DUL) list nor the lists with recent spammers, therefore this study provides statistics on the data without those two lists.	<a href="#">link</a>
<b>UCEPROTECT - Level 1</b>	IP addresses with wrong, missing or generic reverse DNS (PTR record), “dialup” connections, computers with exploitable security holes as well as addresses which are as-	<a href="#">link</a>

<sup>61</sup> See <http://samba.anu.edu.au/rsync/> for more information on `rsync`.

<sup>62</sup> See <http://www.gnu.org/software/wget/> for more information on `wget`.

<sup>63</sup> See <http://spamlinks.net/filter-dnsbl-lists.htm> for a complete list.

<sup>64</sup> See <http://www.corpit.ru/mjt/rblDNSD.html#zff> for a format description.

	signed to well-known spammers.	
<b>dnsbl.ahbl.org</b>	Broad range of listings including open proxies, open relays, well-known spam sources, formmail spam, spam supporters etc.	<a href="#">link</a>
<b>dnsbl.njabl.org</b>	Lists open relays, proxies, dial-up respectively dynamic IP addresses as well as systems that directly send spam.	<a href="#">link</a>
<b>CBL</b>	Lists single IP addresses exhibiting characteristics which are specific to open proxies which have been abused to send spam, worms/viruses or some types of trojan-horse. The CBL collects recent spam activities from more hundreds of mail server streams and tens of thousands spam-traps in at least four continents.	<a href="#">link</a>
<b>NiX spam</b>	Lists of single IP addresses used by recent spam senders, automatic delisting after four days.	<a href="#">link</a>
<b>Spamhaus lists</b>	Spamhaus lists are categorised by multiple types. The SBL is a realtime database of IP addresses of verified spam sources and spam operations. The XBL is a realtime database of IP addresses of illegal 3rd party exploits. The PBL is a database of end-user IP address ranges which should not be delivering unauthenticated SMTP email.	<a href="#">link</a>
<b>Bogon ranges</b>	The bogon ranges describe IP blocks not allocated by IANA and RIRs to ISPs and organisations, plus net ranges reserved for special use by RFCs. The data was taken from completewhois.com.	<a href="#">link</a>
<b>dnswl.org</b>	Whitelist of known legitimate email servers to reduce the chances of false positives while spam filtering, split into four trust levels. Only data with at least low trust level <sup>65</sup> was considered.	<a href="#">link</a>

The mapping between IP and Autonomous System (AS) was made with the help of data published by potaroo.net<sup>66</sup>. The mapping between IP and country was made with the help of data published by completewhois.com<sup>67</sup>. The mapping between IP and RIR was made with the help of data published by the IANA<sup>68</sup>. The mapping between AS number and AS name was made with the help of data published by cidr-report.org<sup>69</sup>.

<sup>65</sup> The trust level of dnswl.org describes the likelihood to receive spam from a listed source. Entries with a trust level of 'None' were not considered.

<sup>66</sup> See <http://www.potaroo.net> for more information.

<sup>67</sup> See <http://www.completewhois.com/statistics/data/ips-bycountry/rirstats/> for more information.

<sup>68</sup> See <http://www.iana.org/assignments/ipv4-address-space> for more information.

<sup>69</sup> See <http://www.cidr-report.org/as2.0/autnums.html> for the complete mapping.

### 7.3 Assessment of blacklists

Currently there is no standardised process of assessing blacklists. Since blacklists are a highly sensitive anti-spam method, the choice for a blacklist should be done very carefully. However, first approaches to give standards on blacklists have been done. The IRTF's Anti-Spam Research Group (ASRG)<sup>70</sup> develops two documents related to DNS based lists such as DNSBLs or DNSWLs. An RFC draft by J. Levine et al<sup>71</sup> describes the structure and usage of DNS based lists and gives a basis for how DNS blacklists work. A draft by Y. Shafranovich, N. Nicholas et al<sup>72</sup> tries to give guidelines for managing DNS based lists and seeks BCP status.

Both papers concentrate on practices for list operators. On the other hand there is less support for users of black- or whitelists. Choosing such a list is not a simple issue. Often it is recommended either to spend a lot of time in evaluating the best lists or taking expertise by some professionals. The second mentioned draft suggests email administrators to consider well which blacklist to pick:

*It is the responsibility of the system administrators who adopt one or more DNSBLs to evaluate, understand, and make a determination of which DNSBLs are appropriate for the sites they administer. If a system or network administrator allows a third party to make blocking decisions for its network, then the administrator MUST understand the policies and practices of those third parties because responsibility for blocking decisions remain ultimately with the administrator.*

A blacklist cannot be classified as good or bad. Several use cases define specific requirements, and in each individual case a review of existing blacklists should be done. Many factors can be reviewed in order to use the most suitable blacklists:

- Performance indicators (false positive rate, true positive rate, activity of the list)
- Way of input (automated, semi-automated, manually, check against whitelist)
- Type of output (via rsync or http, only DNS, only zone transfer)
- Lifetime of entries (fixed or dynamic lifetime, no expiration, escalation)
- Intended use of the list (blocking, scoring, to fan fear)
- Listing policies describe the types of list entries
- Removal process (none, manual, for money, automated)
- Scope of listings (single addresses, net ranges)
- Probing of entries (all, only samples, none)
- Organisational matters (history, details on operators, available contacts)
- Known public reputation
- Costs of usage
- Geographically biases

Not all factors are equally important. For an assessment of the quality and efficiency of a blacklist a small set of those can be taken into account. Picking the most important factors and creating a radar graph helps finding out, if a blacklist suits the individual needs.

---

<sup>70</sup> See <http://asrg.sp.am/>

<sup>71</sup> See <http://tools.ietf.org/id/draft-irtf-asrg-dnsbl-02.txt>

<sup>72</sup> See <http://www.nabble.com/DNSBL-BCP-v.2.0-t3196169.html>



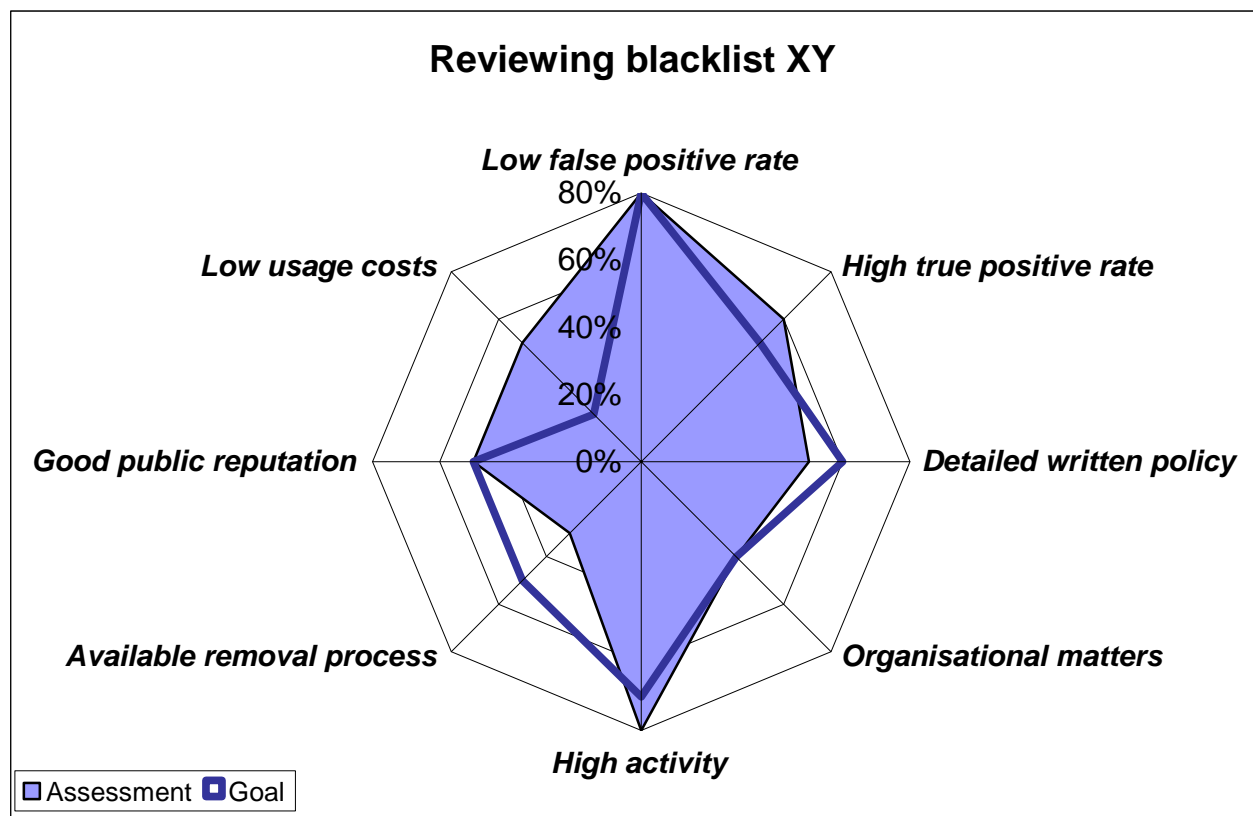


Figure 12: This radar graph describes if blacklist XY fits the needs

The blue line in this graph describes the intended qualities of the blacklist. The higher the percentage, the higher are the expectations from a blacklist. The purple area describes the actual status of the blacklist, i.e. how well it manages the criteria. If the area exceeds the line, the blacklist is beyond the expectations. White gaps between the line and area show that the blacklist is in lack of quality. Summarising the graph shows, whether the blacklist can be used or not. The more white gaps within the circle of the goal line can be found, the less the blacklist suits to the individual needs.

In this case, reviewing the blacklist XY, it outreaches especially the expectations on low usage costs. On the other hand it does not have a satisfying removal process as well as insufficient written policies. Calculating the risk and comparisons with other blacklists might lead to a decision, which blacklist should be taken. In this case blacklist XY could be recommended as an input for a scoring system. Taking a low risk it might be used also for blocking purposes. Comparing this radar graph with graphs of other blacklists would help to figure out the most suitable blacklist(s).

Future research will be done to establish a guided wizard providing individual radar graphs as described before. Therefore for each of the criteria listed above an assessment of blacklists must be done.

## 7.4 Coverage of blacklists

### 7.4.1 Reputation of the IP address space

As of 2007 email communication runs on IP version 4 (IPv4), providing theoretically 4,294,967,296 ( $2^{32}$ ) possible IP addresses of senders. After having subtracted reserved nets an amount of merely 3,706,650,624 IP addresses remains<sup>73</sup>.

In contrast, the amount of IPv4 addresses assigned to Autonomous Systems is more accurate. This comprises 2,968,251,898 addresses as of making the following calculations<sup>74</sup>. However, the best figure for calculations about the coverage of blacklists is the amount of assigned and via a routing protocol advertised IP addresses, which were 1,741,609,238 when making the following calculations.

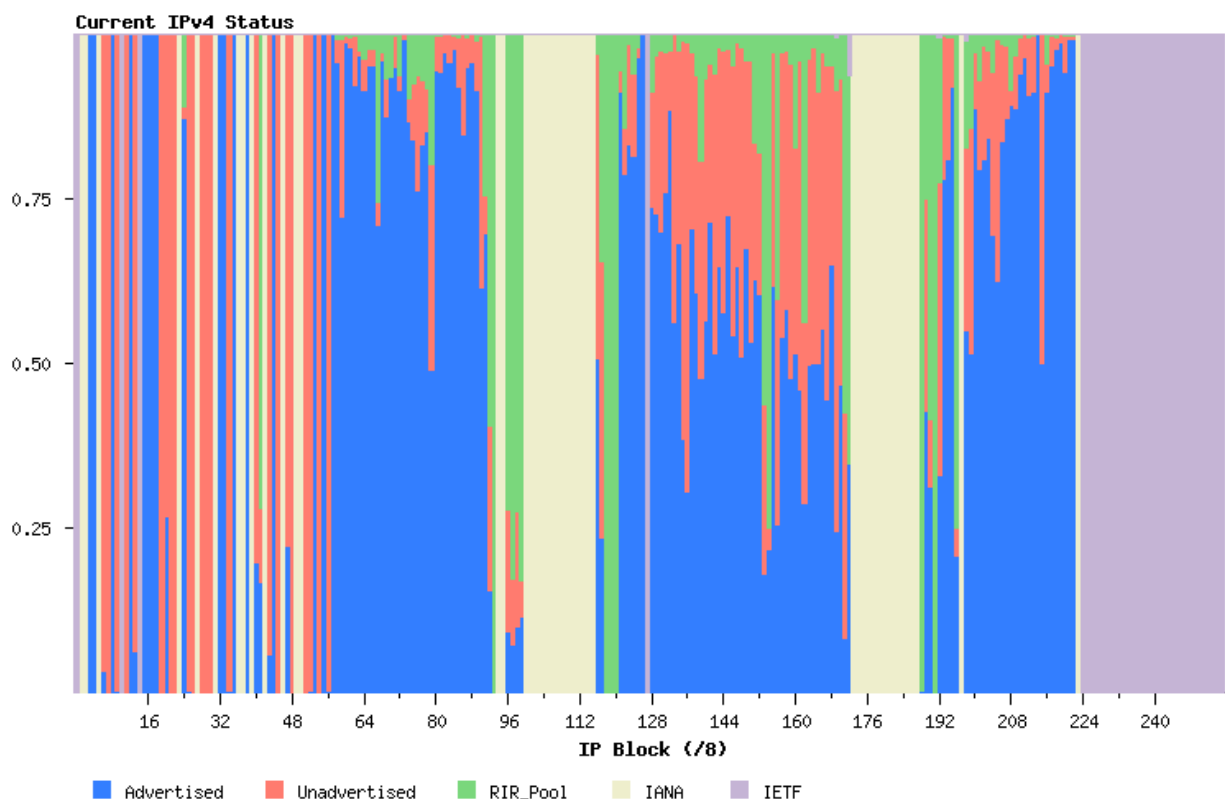


Figure 13: Current IPv4 status shown as kind of number line

This drawing published by IANA visualises the status of the IPv4 address space. On the x-axis the IPv4 address space is printed, each /8 net containing 16,777,216 addresses (e.g. 16.0.0.0 - 16.255.255.255)<sup>75</sup> fills one vertical line. Usable and routable net ranges are the advertised blue stripes. Red areas show assigned, but not advertised and not routable addresses. Free address space that is contained within the bogon ranges is displayed coloured yellow (managed by IANA) or green (managed by RIRs). The remaining grey parts describe IP address space that has been reserved by the IETF.

<sup>73</sup> See <http://www.bgpexpert.com/addressespercountry.php> for more information.

<sup>74</sup> See <http://bgp.potaroo.net/ipv4-stats/allocated-all.html> for the current status of IPv4.

<sup>75</sup> See <http://public.pacbell.net/dedicated/cidr.html> for a description of CIDR.

A union of all available blacklists<sup>79</sup> gives an idea about the part of the IPv4 address space, on which in a way an opinion can be formed. All in all about 18.576% of the advertised net ranges are covered by any kind of blacklist, which equals 323,516,606 IP addresses. Moreover about 0.165% of IP addresses are covered by the whitelist dnswl.org. The remaining 81.26% have no reputation on the Internet represented by entries in the most famous public black- or whitelists. Concluding more than four out of five potential email senders cannot be assessed as good or bad by reviewing only an IP address, showing a big potential for both, spammers as well as black- and whitelists.

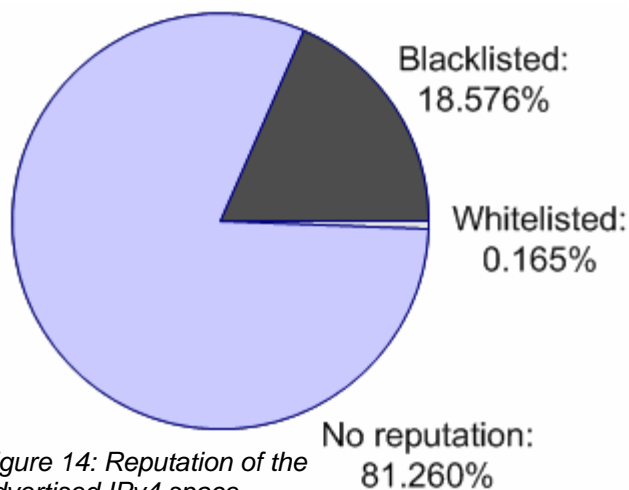


Figure 14: Reputation of the advertised IPv4 space.

A logarithmic presentation shows similarities between blacklists. The higher the black area, the more IP addresses are listed in this part of the IPv4 address space. Each vertical pixel represents a /9 network, i.e. 8,388,608 single IP addresses. Visualising the union of all available blacklists gives a rough view on the IPv4 space with bad reputation.

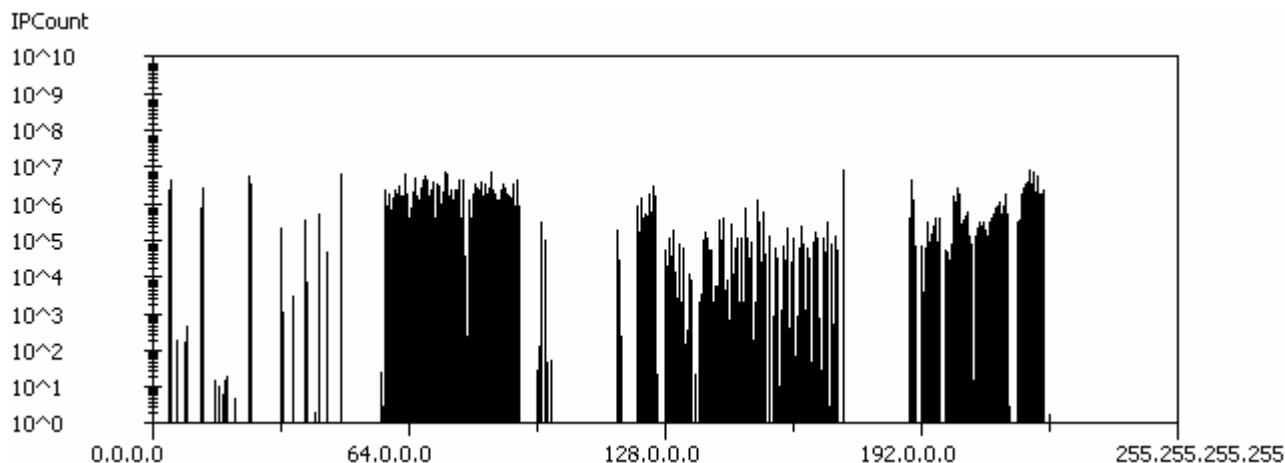


Figure 15: Graphical view of the union of all blacklists

Graphical views on the coverage of each available blacklist<sup>79</sup> reviewing the IPv4 address range are available in annex F.

### 7.4.2 Status of each blacklist

Each black-/whitelist alone covers a percentage of the used IPv4 address space in use. The following table gives detailed information for each list:

Table 3: Coverage of blacklists

Name	Covered range	# of entries	% of IPv4	% of advert.
Bogon ranges	1,413,330,710	7,505	32,9067%	n/a <sup>76</sup>
pbl.spamhaus.org	320,152,555	130,957	7.4541%	18.3826%
xbl.spamhaus.org	5,789,717	5,789,717	0.1348%	0.3324%
CBL	5,212,806	5,212,806	0,1214%	0.2993%
all.dnsbl.sorbs.net	5,090,338	2,836,101	0,1185%	0.2923%
dnsbl.njabl.org	4,459,656	4,459,656	0,1038%	0.2561%
dnsbl.ahbl.org	3,488,407	3,132,255	0,0812%	0.2003%
dnswl.org	2,867,573	9,258	0,0668%	0.1647%
sbl.spamhaus.org	1,807,939	5,222	0.0421%	0.1038%
UCEPROTECT - Level 1	801,283	801,283	0,0187%	0.0460%
NiX Spam	78,677	78,677	0,0018%	0.0045%

The data is ordered by descending coverage of the entire advertised IPv4 address space. This address space has been announced by Autonomous Systems and is routable to a destination. All other addresses are either not advertised or for private use. The last column of the table gives indications about how much of the used IP address space is covered by the blacklists.

Leading with a third of the entire IPv4 address space are the bogon net ranges, including all private and not yet assigned networks. The pbl.spamhaus.org is a list consisting of many big listed networks. It is by far the biggest available blacklist and covers 18.4% of the advertised IPv4 address space. Usually other blocking lists cover between 0.2% and 0.34% of the advertised IP address space, only sbl.spamhaus.org and UCEPROTECT - Level 1 have an even lower coverage of 0.11% and 0.05%.

Two special cases are the whitelist dnswl.org and NiX Spam. Dnswl.org lists 0.07% of the whole Internet, but has with 9,258 entries a very little coverage of all cases needing a whitelisting. Every entry usually belongs to an organisation, so this list has still big potential to improve in the future. Another outlier is NiX Spam, covering only a very small part of the Internet. NiX Spam is based on their philosophy to list entries only with a very short lifetime of four days. In comparison to a similar blacklist such as the CBL, NiX Spam is a very small list (factor 66) and therefore only a small part of recent spam sources is covered. Nevertheless it has a hit rate of above 20%, showing the high concentration of recent spam sources.

### 7.4.3 Potential for lists to block botnets

Spamhaus' blacklist pbl.spamhaus.org (PBL) covers dynamic IP address spaces announced by providers or added by Spamhaus manually, which are not supposed to send unauthenticated SMTP email. Although covering 320 million IP addresses, there might be

<sup>76</sup> Bogon ranges are per se not advertised, therefore this number would be senseless.

a bigger potential for the Spamhaus PBL. The real address space allocated to dialup- and broadband users is probably bigger due to the high amount of people online<sup>77</sup>.

Most spam comes via botnets and should be covered completely by the PBL. The fact that this is currently not the case is revealed by a comparison with the CBL. The CBL lists recent spam sources, which are likely to be bots. Subtracting all networks listed in the PBL from the CBL shows a remaining set of IP addresses, which attracted attention by sending spam and were not announced as such sources in the PBL before.

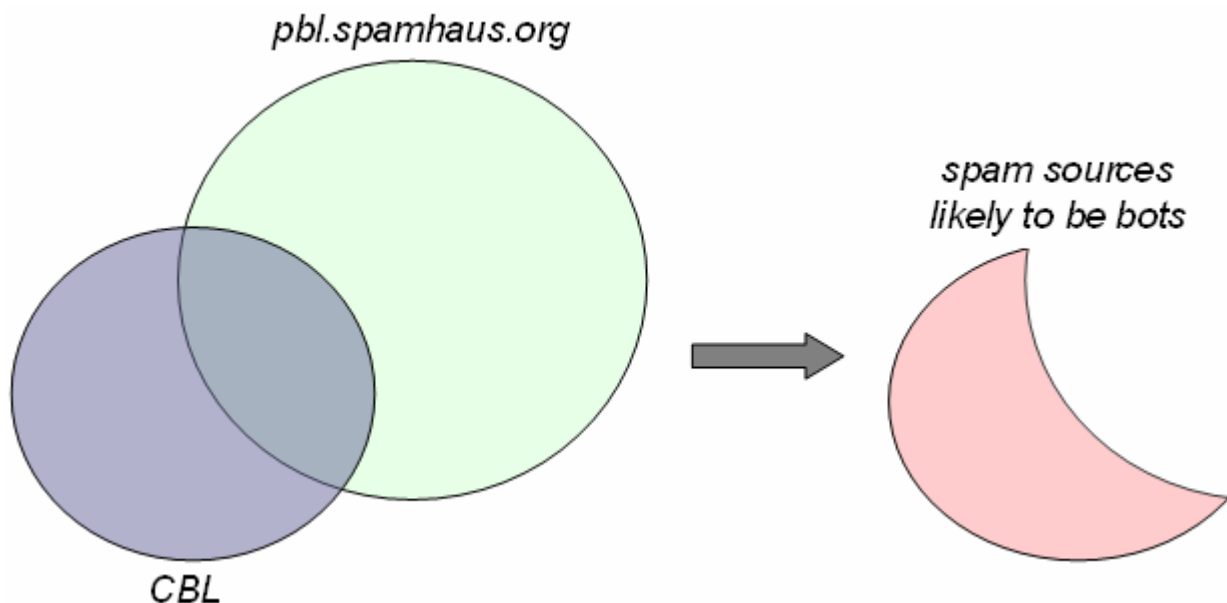


Figure 16: Finding out the potential of a list blocking botnets

When building this set of IP addresses after all 27.2% of the range covered by the CBL remains. These 1.42 million entries are only a small part of the near-threatened address space. They show that the PBL is by far incomplete with its coverage of less than a fifth of the advertised IPv4 address space.

In order to mitigate this issue, Spamhaus should expand its actions for countries where they do not have a broad coverage. Improving their business especially in China would increase the coverage of the PBL a lot. However, providers signed on to Spamhaus often do not have precise plans of their network. In the consequence they cannot publish complete information and report only parts of the actually applicable networks.

But the coverage of the blacklists is by no means an indication for the quality of a blacklist. Old listings or listings of unused IP address space could push this number without any positive consequence on the efficiency. On the other hand a small list with the worst spam sources is likely to be very efficient. Therefore the list pbl.spamhaus.org does a good job, equally whether it is complete or not.

<sup>77</sup> The statistics on <http://www.internetworldstats.com/stats.htm> show 1,154 million people using the Internet, which is approximately factor 3.57 to the listed dialup space. Although this is not a one-to-one mapping, since users can share a dialup slot, 3.57 users on one dialup IP would be very weak facilities and is not realistic.

#### 7.4.4 Recent vs. old spam sources

Comparing the listings of only recent spammers with more static blacklists is very interesting. It shows whether a part of the IPv4 address space was used by spammers recently. The following graph shows a comparison between the NiX Spam and pbl.spamhaus.org blacklists. NiX Spam lists spammers for a short period of four days and thus includes only recent spam sources. On the other hand pbl.spamhaus.org is a very static list containing many end-user IP addresses.

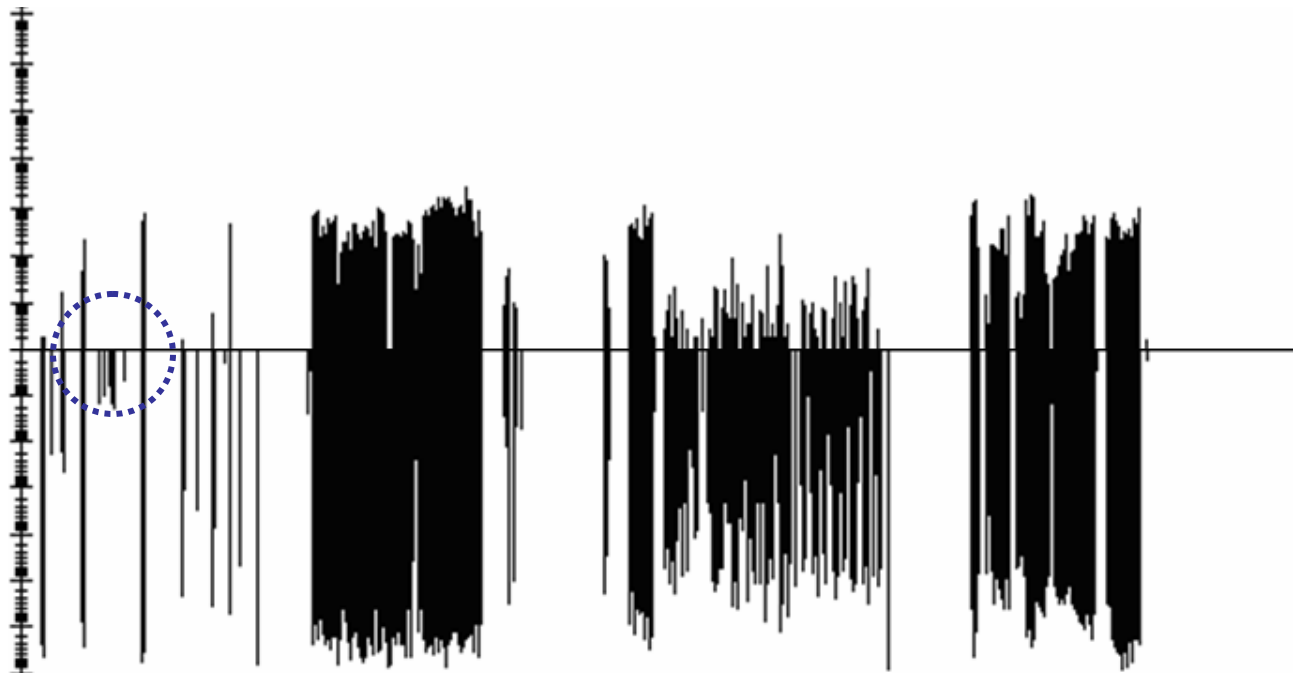


Figure 17: Graphical comparison between NiX Spam (top) and pbl.spamhaus.org (bottom)

The graph shows the IPv4 address space, split symmetrically with the horizontal line in the middle. In the upper part of the graph all entries of the NiX Spam list were drawn, the lower part represents the entries of pbl.spamhaus.org. Some of networks blocked by pbl.spamhaus.org are not listed in the NiX Spam list<sup>78</sup>, i.e. they were not used for spam waves four days prior the data snapshot.

### 7.5 Blacklist entries by country

#### 7.5.1 Regional biases of blacklists

Assigning each blacklist entry to a country shows how many clients from a country send respectively have sent spam. This might give ideas about whether a blacklist operates locally or has its emphasis in specific countries. An assignment of each available blacklist to countries is available in annex D.

In five of nine blacklists the United States of America lead the table of listed hosts, China in further three cases and Brazil in the blacklist CBL. China and the U.S. are in every blacklist in the top-4 positions, what might be caused by the big amount of Internet users in these countries. European countries are rarely within the top-5 countries, only France, Poland and Germany are particularly mentioned there.

<sup>78</sup> See the blue circle on the left of the graph for a marking.

The whitelist dnswl.org seems to have strong biases towards middle Europa. Austria and Switzerland are ranked with a much higher ranking than in other lists. This might be caused by the Austrian operator of this list, currently working at a Swiss employer. Reviewing the locality of lists especially NiX Spam with a special focus on Germany and UCEPROTECT - Level 1 with a special focus on Poland attracted attention.

Of special interest is the blacklist pbl.spamhaus.org. Providers publish their net ranges used for dynamic hosts in this list and Spamhaus adds manually net ranges if providers do not. These hosts should not send emails without relaying via smarthosts and can therefore be blacklisted. If the ratio between IP addresses assigned to a country and listed addresses is high, the providers in this country do a good job in keeping transparency of their networks. Approximately more than a half of all assigned IP addresses are used by dynamic hosts, therefore a quota of at least 50% for each country should be intended. Since the average quota of the top-50 countries is only 21%, there is high potential in collecting IP address blocks of dynamic hosts.

### 7.5.2 Union of all blacklists

The following table shows the union of all available blacklists<sup>79</sup> assigned to countries. *Entries* describes the amount of blacklistings assigned to a country, *range* indicates the total amount of IP addresses which are assigned to that country and coincidentally blacklisted. *Quota* gives information about the percentage of the covered range of a country, i.e. which ratio of the amount of a country's assigned IP addresses is listed in a blacklist. Grey rows are countries from the EU.

Table 4: Countries with the most blacklisted IP addresses

rank	country	entries	range	quota
1	United States	283463	121389419	8.72%
2	Japan	54708	29054339	18.81%
3	China	238876	27655748	22.88%
4	Germany	23064	23590123	34.20%
5	(unknown)	8599	17304975	n/a
6	Canada	40958	10656784	14.60%
7	United Kingdom	42859	7819107	9.47%
8	France	43123	6982290	10.88%
9	Taiwan (Province Of China)	16044	6931085	37.00%
10	Mexico	35877	6352605	39.07%

### 7.5.3 Discredited countries

Since a main part of the entire union consists of pbl.spamhaus.org, a high rank in this list is not an indication for a spammer friendly country<sup>80</sup>. Another more accurate indication is a low quota. On the one hand this might be an indication about proper networks in a country. On the other hand it could be low because many providers did not publish their dynamic host ranges properly. However, this approach is not satisfying accurate either. Therefore a closer look to known but not as such declared spam sources helps finding precise data.

The CBL is a list with recent spam sources with a very good reputation. Spamhaus includes this data in their XBL list. Moreover Spamhaus manages the PBL, which consists of

<sup>79</sup> See chapter 7.2 for a complete list of the available blacklists.

<sup>80</sup> Providers publish their IP address ranges used for dynamic hosts in this list. These hosts should not send emails without relaying via smarthosts and can therefore be blacklisted.

voluntary listings made by providers about ranges, that shouldn't be used for unauthenticated SMTP traffic. When subtracting all IP addresses listed in the PBL list from the CBL list, the remaining part shows recent spam sources that were not declared as such by providers. Therefore the result shows networks, which are likely to be used by spammers without counter-measures of the providers.

Table 5: Countries with recently spam sources not listed in pbl.spamhaus.org

rank	country	CBL	CBL-PBL	CBL remaining
1	China	650290	231290	35.57%
2	Brazil	680678	166166	24.41%
3	Korea, Republic of (South)	211228	163311	77.32%
4	United States	388390	137678	35.45%
5	Russian Federation	148319	54561	36.79%
6	United Kingdom	84331	29063	34.46%
7	France	115280	27960	24.25%
8	Argentina	104760	25898	24.72%
9	Peru	48435	25366	52.37%
10	India	288320	23611	8.19%

The table shows the results of subtracting pbl.spamhaus.org from the CBL list. It is ordered by descending remaining entries (*CBL-PBL*) after subtraction. Percentages in the column *CBL remaining* indicate, which part of the CBL remains after deleting all intersections with the PBL list.

Only two European countries are in the top-10 of this special set of spamming sources. When comparing with the union before, Germany has left the top-10 and has a good coverage by Spamhaus' PBL list. Only 6.45% of German CBL entries remain when subtracting the PBL from it. On the other hand, France (24.25%) and especially United Kingdom (34.46%) have bad ratios of remaining CBL entries. This leads to a high listing of spam sources that were not advertised in the PBL.

These data show the potential of a list like the PBL. Most of the spam is coming from end-user computers which are not supposed to send messages via unauthenticated SMTP. Providers should start to publish their address space assigned to such users.



## 7.6 Blacklist entries by AS

Some Autonomous Systems do not protect their networks as they should do and therefore have a high amount of spam originating from those. Grouping a blacklist's entries by Autonomous Systems might be a good evidence to get a reputation of it. This data is available in detail for each blacklist<sup>79</sup> in annex E.

When reviewing some blacklist specific data, the practices of blacklist operators become visible. Some Autonomous Systems are completely listed in blacklists, to name a few AS 11784 at all.dnsbl.sorbs.net or AS 35935 at sbl.spamhaus.org. Very interesting is a closed look to Spamhaus' PBL listing the dialup-/broadband IP address ranges announced by providers, because it shows whether providers submit this kind of information. With Deutsche Telekom AG, Telecom Italia and France Telecom the three biggest ISPs in Europe announced information to this list.

Regarding the whitelist dnswl.org it is questionable, if Autonomous Systems need listing with net ranges covering 65,536 IP addresses. Usually email servers within huge networks should be concentrated on smaller net blocks. Moreover it seems very obvious, that dnswl.org has good connection to Switzerland, since about 20% of the top-50 listings are from the Alps country.

Big fishes regarding Autonomous Systems are AS 9121 (TTnet), AS 4837 (China169 Backbone), 3320 (DTAG) and 5617 (TPnet.pl), appearing very often within the top-5 listed systems. In general the quota<sup>81</sup> of an AS rarely exceeds one percent, showing the incompleteness of most blacklists. Only Spamhaus' PBL lists reasonable bigger amounts of addresses, covering two-digit quotas.

The following table is based on a union of all available blacklists<sup>79</sup>. It is ordered by the summed size of all blacklist entries by AS, showing the top-10 of all Autonomous Systems which have the most listings in the union of all blacklists. A top-50 list is in Annex E.

Table 6: Autonomous Systems with the most blacklisted IP addresses

rank	asid	name	entries	range	quota
1	17676	JPNIC-JP-ASN-BLOCK Japan NIC	217	20448878	27.85%
2	3320	DTAG Deutsche Telekom AG	893	15395695	62.04%
3	3356	LEVEL3 Level 3 Communications	4952	9997881	8.60%
4	209	ASN-QWEST - Qwest	7128	5553678	3.65%
5	1668	AOL-ATDN - AOL Transit Data Network	14	5374728	38.55%
6	7018	ATT-INTERNET4 - AT&T WorldNet Services	8812	5137125	2.77%
7	4837	CHINA169-BACKBONE CNCGROUP Backbone	79476	4561014	22.36%
8	5089	NTL NTL Group Limited	3591	4053737	33.94%
9	7132	SBIS-AS - AT&T Internet Services	39005	3438800	11.24%
10	5430	FREENETDE freenet Cityline GmbH	815	3149354	94.22%

### 7.6.1 Discredited Autonomous Systems

Since a main part of the union consists of pbl.spamhaus.org, a high rank in this list is not an indication for a spammer friendly network. On the contrary, the top listed providers are likely to do something against spamming, as far as most of their listings origin from pbl.spamhaus.org. However, protecting their network from outgoing spam via other meth-

<sup>81</sup> For a description of the quota please the annex E.

ods like managing port 25 is more efficient and saves resources of both sender and recipient.

When removing all intersections from the CBL list with pbl.spamhaus.org, only spam sources that the provider did not declared as such remain. Most spam is sent via botnets, i.e. dialup hosts that should be listed in the PBL list. The following table gives an idea about networks, which have probably not announced their complete end-user dialup ranges to Spamhaus.

Table 7: Autonomous Systems with recently spam sources not listed in pbl.spamhaus.org

rank	asid	name	entries	range	quota
1	4837	CHINA169-BACKBONE CNCGROUP Backbone	89465	89465	0.44%
2	8167	TELESC - Telecomunicacoes de Santa Catarina SA	80844	80844	3.77%
3	4766	KIXS-AS-KR Korea Telecom	51513	51513	0.24%
4	3215	AS3215 France Telecom - Orange	17728	17728	0.18%
5	9121	TTNET TTnet Autonomous System	16921	16921	0.17%
6	7132	SBIS-AS - AT&T Internet Services	13835	13835	0.05%
7	15475	NOL	11216	11216	1.61%
8	4230	Embratel	10849	10849	0.30%
9	4808	CHINA169-BJ CNCGROUP Beijing Province Network	10249	10249	0.21%
10	27699	TELECOMUNICACOES DE SAO PAULO S/A	10151	10151	0.42%

France Telecom is the highest ranked European provider. Almost every 500<sup>th</sup> IP address announced by them is listed in the CBL and not in pbl.spamhaus.org. Moreover France Telecom announced only 18.9% of its routable IP addresses to the PBL list. This shows, as one example of many, the big potential of the PBL list.

## 7.7 Intersections between blacklists

Usually blacklists have similar ways how to get new data into their spammer database. Moreover volunteers often contribute to multiple blacklists or administrators of blacklists help each other go get better data. For this reason it is not very surprising that blacklists amongst themselves have some intersections, i.e. that an IP address listed in blacklist A is also listed in blacklist B.

The following table gives the percentages about which amount of IP addresses listed in blacklist A (row) is covered by blacklist B (column). The table's data is asymmetric, because the sizes of the blacklist differ. In other words, blacklist A covering 100 IP addresses can be covered with 8% by blacklist B with 10.000 entries, whereas vice versa blacklist B is only covered with 0.08% by blacklist A.

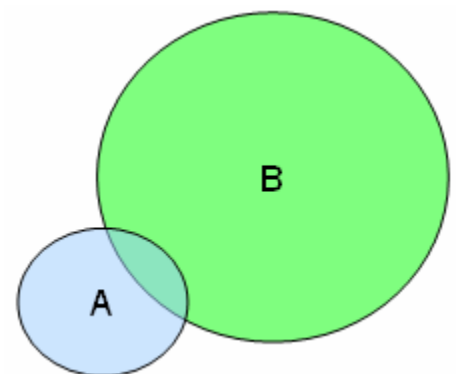


Figure 18: Example of intersections

High percentages are coloured differently. If a blacklist is covered more than 70% by another blacklist, it is coloured red. The colour yellow indicates intersections between 40% and 70%, whereas light green stresses percentages higher than 10%.

reference \ comparison	all.dnsbl.sorbs.net	UCEPROTECT L1	NiX Spam	dnsbl.ahbl.org	sbl.spamhaus.org	dnsbl.njabl.org	CBL	pbl.spamhaus.org	xbl.spamhaus.org	dnswl.org	Bogus ranges
all.dnsbl.sorbs.net	-	1,83	0,28	10,17	10,67	11,03	8,03	36,92	17,92	0,002	7,73
UCEPROTECT L1	11,61	-	2,34	1,97	0,58	2,93	64,14	69,96	64,79	0,026	0,01
NiX Spam	18,32	23,80	-	1,79	0,64	2,58	41,02	55,36	42,58	0,064	0,02
dnsbl.ahbl.org	14,83	0,45	0,04	-	0,56	64,32	3,74	66,38	13,87	0,002	0,22
sbl.spamhaus.org	29,15	0,25	0,03	1,04	-	0,88	1,23	5,49	1,49	0,003	9,68
dnsbl.njabl.org	12,59	0,53	0,05	50,31	0,37	-	4,75	67,11	21,03	0,003	0,28
CBL	7,84	9,86	0,62	2,50	0,44	4,07	-	73,91	100,00	0,001	0,00
pbl.spamhaus.org	0,58	0,17	0,01	0,72	0,03	0,93	1,19	-	1,36	0,000	1,48
xbl.spamhaus.org	15,39	8,76	0,57	8,17	0,47	15,83	88,05	73,92	-	0,001	0,01
dnswl.org	0,003	0,007	0,002	0,003	0,002	0,005	0,001	0,002	0,002	-	0,027
Bogus ranges	0,03	0,00	0,00	0,00	0,01	0,00	0,00	0,34	0,00	0,000	-

Figure 19: Intersection matrix between IP based black- and whitelists

Discussing the red markings makes dependencies between the different Spamhaus lists visible. Obviously the XBL integrates the complete CBL, and a big part (74%) of the dynamically listing CBL is also integrated in the more or less static PBL that lists net ranges. As a conclusion the XBL has a reasonable coverage of 74% by the PBL, so the Spamhaus' lists are by no means completely distinct.

Logically the bigger blacklists (pbl.spamhaus.org is the extreme) are covered less than smaller lists. In the opposite, the smaller lists UCEPROTECT - Level 1 and NiX Spam are covered with main parts by the Spamhaus lists. Taking the columns of the table into account one can see how much a blacklist covers other lists. For instance all.dnsbl.sorbs.net has many intersections with other lists, covering 10%-30% in six cases.

As mentioned earlier the blacklists NiX Spam and CBL are very similar in their way of data input. Both blacklists use spamtraps and traffic analyses to save recent spam activities automated in a database. Whereas 74% of the CBL listings hit the end-user IP addresses of the PBL, NiX Spam has a lower intersection with only 55%. NiX Spam seems to cover a higher percentage of static spam hosts (like open relays, open proxies etc.) than the CBL. On the other hand the CBL contains many end-user addresses, which might be caused by the high percentage of spam coming from botnets hosted at end-user PCs.

The whitelist dnswl.org can be reviewed with special interest. Because the intersections between black- and whitelists should be very low, the highlighting of the percentages is more sensitive. Both UCEPROTECT - Level 1 and especially NiX Spam show high ratios of intersections with 50 respectively 210 IP addresses. This might be caused either by the

way of integrating data into the blacklist (maybe too aggressive) or by the locality of the two blacklists and the whitelist (all three are operated from Central Europe).

In general the intersection between white- and blacklists shows big issues of blacklisting. Deciding whether an email server should be listed is problematic. Many servers often send spam as well as ham. For instance freemail providers offer services against a cost-free registration. Although this registration is secured by CAPTCHA mechanisms, some spammers trick them and send spam via (for the rest) legitimate email servers. Another issue are spammers stealing a person's email credentials and sending spam via regular smarthosts. It is up to every black- respectively whitelist-operator how to judge about these cases whether to list or not. The user of blacklisting should be aware of this concern and understand the aggressiveness of the different blacklists.

Finally considering the bogon net ranges it is interesting, that blacklists integrate them only in small parts. Some research could be done to figure out, how often bogon net ranges are used to send spam, e.g. with the help of BGP hijacking. If it was common practice, blocking SMTP traffic from these bogon ranges would decrease spam as well.

The high coverage of bogon entries with dnswl.org might also be contradictory, if the listings are not caused by private LAN addresses or other legitimate sender email addresses. After speaking with the administrator of dnswl.org the intersected entries have been removed from dnswl.org. The admin stated that the five removed entries were typos when inserting them into the list, showing a reasonable risk of this manual processing. It is highly recommended to check against common blacklists before inserting an entry into a whitelist, including lists with bogon net ranges.

## **7.8 Quality assurance for blacklists**

### **7.8.1 Quality indicators for blacklists**

Smattering blacklists have two main indicators of quality. The true positive rate describes how many spam emails get blocked when using a blacklist. On the other hand the false positive rate explains how many legitimate emails (ham) get blocked.

True positive rates are very easy to determine, by implementing spamtraps and analysing incoming email. Therefore many people determine and publish these figures<sup>82</sup>, providing a rough feeling about the efficiency of a blacklist. But only reviewing this data is very dangerous, because the true positive rate itself does not say much about the quality of a list. Cautionary tale is a blacklist blocking the entire Internet, leading to a hit rate of no less than 100%.

For this reason false positive rates of blacklists are of high interest. These rates describe which percentage of legitimate email was wrongly blocked by a blacklist. Similar to deploying a spamtrap, for this case a hamtrap can be implemented. A low false positive rate guarantees the quality of a blacklist and helps meeting the demands of recipients.

### **7.8.2 Measuring false positive rates**

It is much more difficult to receive dedicated ham than dedicated spam. Spammers usually do not do much effort to explore spamtraps, but they spend much time on harvesting of

---

<sup>82</sup> For a list of available statistics see <http://www.spamlinks.net/stats.htm#dnsbls>.

new email addresses. A false positive rate can either be processed manually or in an automated fashion. Unfortunately a manually estimation of false positive rates will not lead to satisfactory results, since the individual classification of spam varies from user to user. Moreover this method relies on human interactions, which cannot be guaranteed with a sufficient amount of data processed.

On the other hand automated methods tend to not represent a typical user's inbox and hence measure everything but realistic percentages. Al Iverson did a first approach measuring ham emails by subscribing to opt-in newsletters. When receiving mail from these bulk senders, Iverson's method inquiries blacklists whether the sender is listed. If the sender is listed on a blacklist, he counts the email as a false positive. Although this is 100% correct, the measured rates do not show a typical user's email receipt. Email recipients usually get more personalised than bulk email and therefore this method covers only an insufficient part of the world-wide email communication.

Since Al Iverson's method was the only automated method measuring false positive, some research has been done as part of this paper in order to develop an improved system. The new system covers as well personalised mails sent by users as well by receiving email from the incoming mailing lists. After subscribing to a mailing list, the hamtrap extracts the senders IP address from the ingoing ham emails. This IP address gets checked against existing blacklists with a similar procedure to Al Iverson's method.

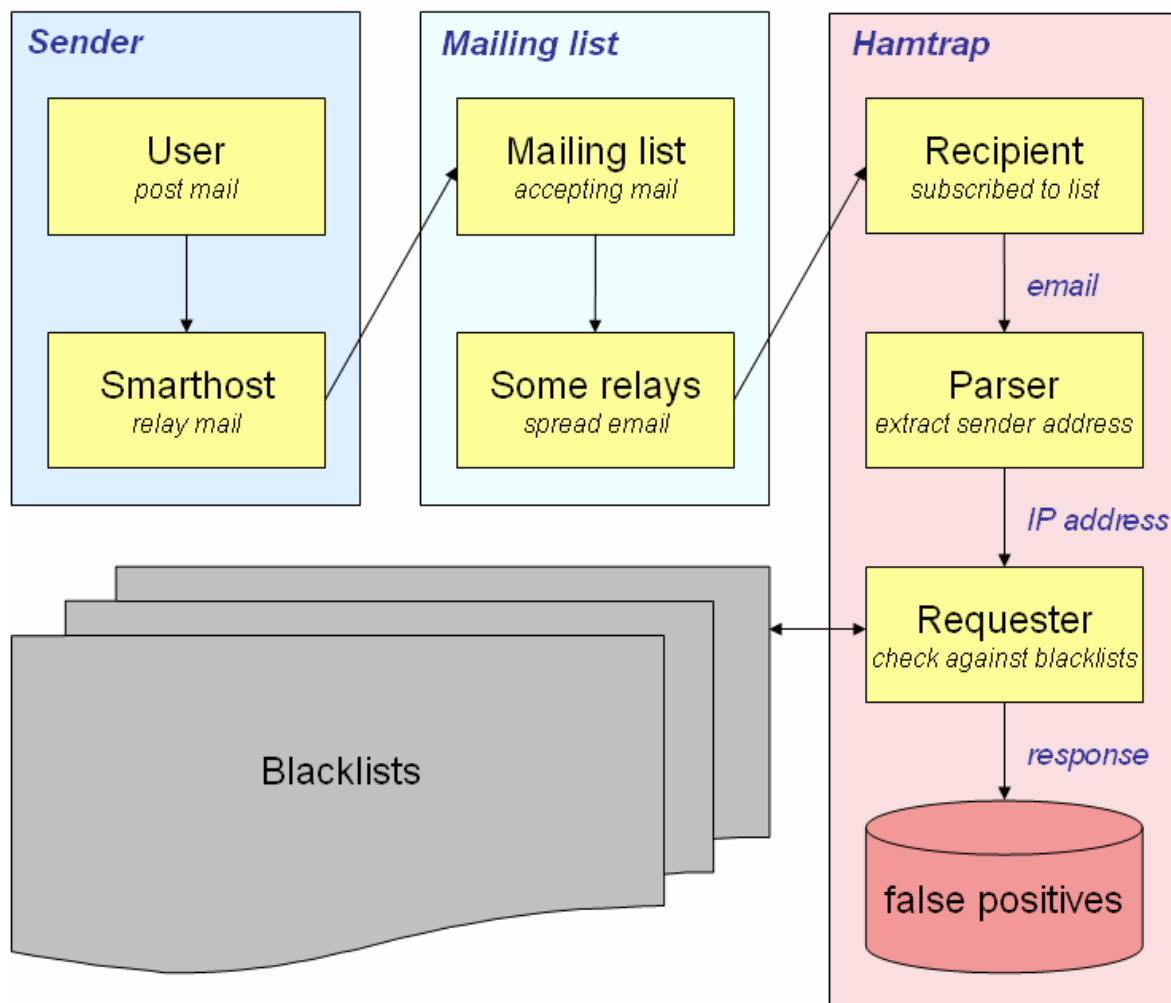


Figure 20: Data flow in the hamtrap installation

The illustration shows the detailed procedure of the developed hamtrap. A mailing-list subscriber usually uses his smarthost in order to send emails to the mailing list. Important is the fact, that the sender uses exactly the same way sending email to the list as when sending email to other people directly. This ensures a good quality of the measurements.

After sending an email to the mailing list, the mailing list spreads the email via relays to all the recipients. One of those recipients is the hamtrap, parsing the senders IP address from the email. This address is simply the address of the server sending the email to the mailing list, in most cases a smarthost. It can be extracted from the received email by analysing the "Received:" of the email's header lines.

After determining the IP address a requester checks it against multiple blacklists. If a blacklist responds positively (i.e. that the sender is listed on it) a false positive has been detected and it is stored to a database. From the ratio between total emails and false positives can be built up a false positive rate by blacklist.<sup>83</sup>

### 7.9 Quality assurance for whitelists

As discussed earlier in chapter 7.7 whitelists usually have intersections between blacklists. This is one of the main reasons why whitelists exist, protecting the recipient against high false positive rates of a blacklist. However, some measures increasing the accuracy and mitigating the risk of whitelists are highly recommended.

When administrating a whitelist, a good tracking system for entries is very useful, making a review of the record's history possible. If the operators receive complaints about listings, they can judge more easily about the case by considering old assessments. Moreover it is suggestive to group whitelist entries in categories and assigning each entry a kind of trust level<sup>84</sup>.

It is quite common that the IP address space is changing over time, since providers change, routing address assignments get withdrawn and announced or machines get assigned to new addresses. Therefore it is mandatory when administrating a blacklist, that the entire list gets checked for changes regularly. One way to do so is comparing current entries with blacklists and reviewing cases with intersections in detail. Although an intersection is not necessary a wrong listing, it indicates a possible conflict.

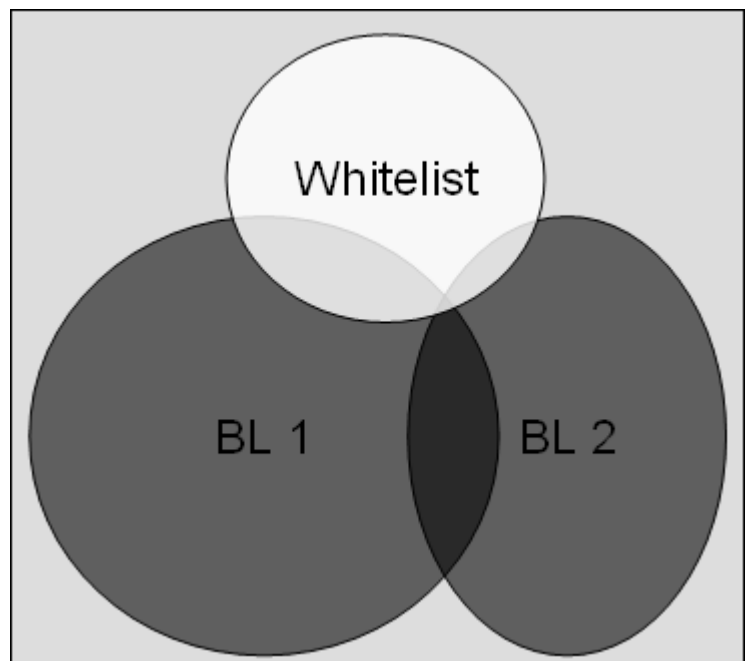


Figure 21: Intersections between black- and whitelists

<sup>83</sup> There are some constraints which have to be valid in order to use the hamtrap as described.

Those have been discussed in the Usenet and can be found at <http://preview.tinyurl.com/224vjw>.

<sup>84</sup> Categorizing is current practice at <http://www.dnswl.org/>, categorizing into four trust levels.

## 8 Anti-spam recommendations for European providers

This chapter gives an overview about current best common practices as well as future oriented practices for European providers. This advice is based on research, the results of ENISA's survey as well as on reviews of many papers related to spam. Furthermore direct contacts to providers were used to exchange information on this topic.

On the Internet many collections of best practices for ISPs or ESPs are available. Nevertheless, the suggestions in this chapter differ from the existing ones. Although available resources may give excellent hints, mainly four reasons make recommendations in this document at least to a very important addition to those existing helps.

1. The results of the European wide ENISA survey helped to get a more international view on the topic of anti-spam measures. Usually the given practices are recommended by national associations, or even by non-EU associations. The latter is especially critical, because non-EU countries can have different legislation on anti-spam. All EU countries have (more or less) the EU Directive 2002/58/EC implemented and similar national legislations regarding electronic communications.
2. The comprehensive research on blacklisting gave lots of empirical information about existing blacklists. This strongly influenced the evaluation of network level blocking methods such as blacklisting. Almost all researched information in this field was never seen on the Internet before and is highly important for assessing blacklisting.
3. Usually advices are very abstract, so that technicians can hardly adopt these to their network without decoding generic texts into explicit technical measures. These recommendations aim to give detailed technical advice in order to assist network operation centres (NOCs) to battle spam.
4. As mentioned in previous chapters, the spammers try to improve their techniques and trick current anti-spam tools. The following recommendations are based on profound and future oriented research in the field of anti-spam and give some long-term strategies to fight spam. Currently existing best common practices might be perfect for the current situation on spam, but these recommendations will help to mitigate the spam problem on the long run.

### 8.1 Highly recommended

#### 8.1.1 Manage port 25

##### 8.1.1.1 General suggestions

As described in chapter 3.1 botnets play a huge role in the business of spam. It is up to every ISP to mitigate this problem in its own network by prohibiting outgoing SMTP connection without proper authentication. If every provider applied these methods properly, in theory the whole botnet spam would be stopped. This seems to be mystic at first sight, since not every provider is really willing to reduce its outgoing spam. However, reducing the amount of spam sent by bots from its network a provider can do its bit in order to mitigate the world wide problem of spam.

Applying this method will decrease incoming spam of the providers only slightly. Nevertheless, for the reason described above and for getting a better reputation from other providers it is highly recommended for providers to manage their port 25 connections. The

effort of doing this is very negligible and if well-planned no disadvantages occur. To evaluate this, looking to a given recommendations by MAAWG is convenient.

Following MAAWG's recommendations<sup>85</sup> on managing port 25, only a few steps are necessary to handle outgoing SMTP connections. In this document these MAAWG recommendations were adopted after small modifications:

1. Block outgoing connections to destination port 25 from all clients on your network, other than those that are explicitly authorised to perform SMTP relay functions.
2. Provide Message Submission for Mail on port 587, as described in RFC 4409, and adopt (if applicable) the MUA software for clients to use Message Submission.
3. Configure, if applicable, email client software to use Message Submission for Mail instead of ancient submission on port 25.
4. Abstain from interfering with outbound and inbound connectivity to port 587.

Usually customers do not need to send mail directly from hosts in a provider's network on port 25 to outside located MTAs. For this reason, blocking outgoing connections on port 25 (see bullet 1) would prevent the unauthenticated submitted spam sent via bots in this network. In order to allow customers to generally use Message Submission for Mail (as described in chapter 5.3.1) the provider's MSA should accept mails both on port 25 and port 587 (see bullet 2).

#### 8.1.1.2 Special cases when blocking port 25

Blocking TCP port 25 usually might interfere with the clients' communications. Some use-cases of communication are imaginable, which would be disturbed by these changes. On the other hand, for each case a solution to manage the occurring issues is conceivable.

- A) It might happen that customers want to deploy an own email server *outside* the providers network. Assuming this server is able to relay mails via SMTP on port 25, the only issue left is email submission to this server which is impossible via SMTP from inside the network. Therefore customers in such situations should be asked too use Message Submission for Mail to submit their mails to this server. In this case it is important for providers to not interfere with outgoing connections on port 587 (see bullet 4).
- B) Especially business customers or some advanced users might want to deploy an own email server *within* the provider's network in order to send emails and therefore do not want to submit their mails to the provider's servers. These servers can no longer relay mails outside the provider's network (after applying port 25 blocking). Two possible solutions are given for this kind of problem:
  - a. The customer's MTA can relay all mails to the provider's MTA, which relays the emails outside the network.
  - b. The provider manages a list of customers that are allowed to send emails outside the network using SMTP on port 25. For simplicity this list could be a

---

<sup>85</sup> English, French and German versions are available at <http://www.maawg.org/port25>



whitelist of IP addresses. In case the customer's IP addresses are dynamic, i.e. they change regularly, this legitimated senders list has to be more advanced (e.g. with a customer's identification instead of an IP address).

For the case clients outside the provider's network have to submit emails to this server within the network, these clients can submit emails using Message Submission for Mail. In this case it is important for providers to not interfere with ingoing connections on port 587 (see 4).

- C) When customers send emails directly to MTAs outside the provider's network using port 25, this happens either based on wrong configuration of the MUA or because the customer's computer was infected and became a spamming bot. In both cases, this kind of usage is undesirable. If customers have problems with sending legitimate emails due to port 25 blocking, they should check the MUA's configuration. The technical support hotline of the provider can help in this case.

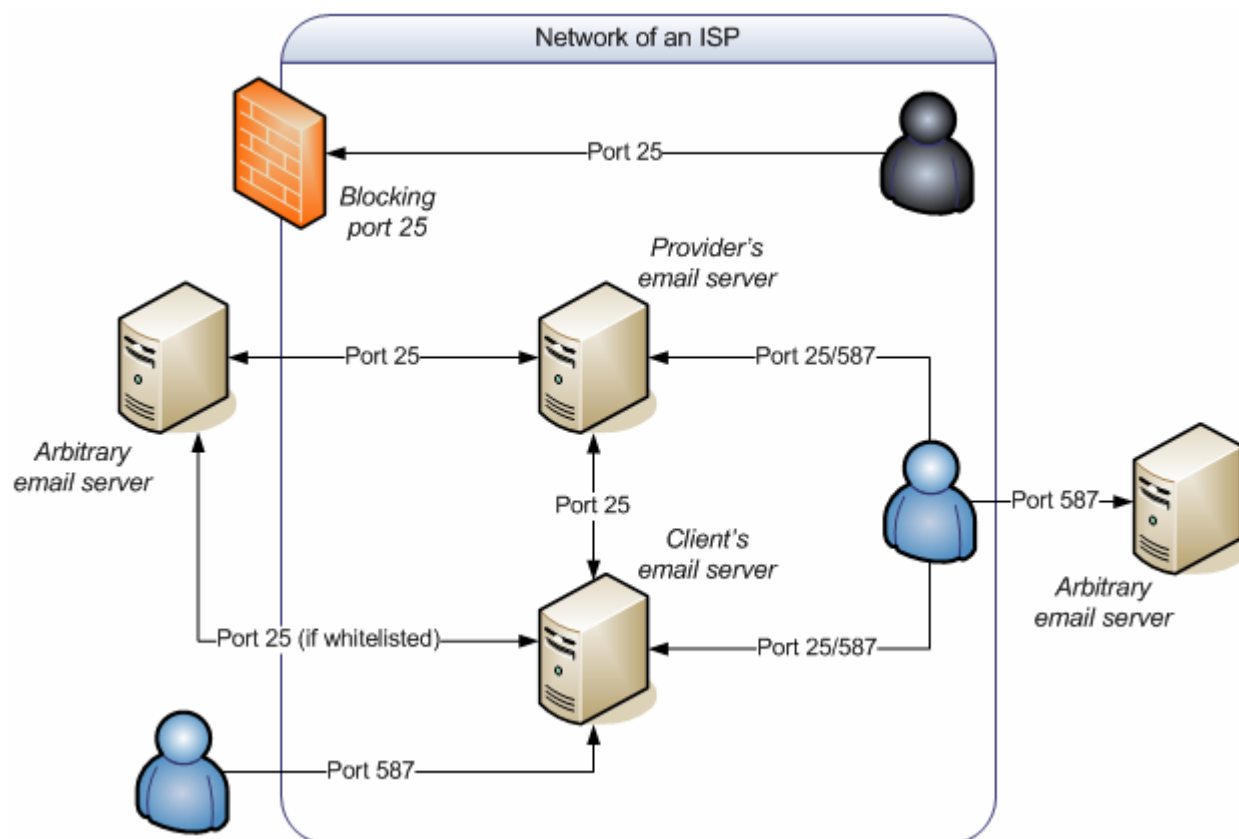


Figure 22: Possible network structure when managing port 25

The illustration shows a possible scenario for each special case mentioned above. A provider's email server can relay and receive emails via port 25. The client's email servers should be whitelisted, if they are authorised to send unauthenticated email. Those special servers are then allowed to send or receive emails on port 25. If they are not authorised, they have to exchange emails via the provider's email server.

#### 8.1.1.3 Handling the transition to blocking port 25

The recommended measures for providers to highly reduce outgoing spam are very easy to apply. Moreover, if communicated to the customers in a good way, usually this ap-

proach does not affect the customers email transfer. However, some ISPs do not want to apply these methods for mostly non-technical reasons. Those providers are solicited to publish their clients' dialup ranges as well as a range of their MTAs in order to allow other providers distinguishing them<sup>86</sup>.

In order to allow an optimal migration to a blocked port 25 the following bullets give a small roadmap and discussions to be considered:

- Debating the situation and its needs with lawyers helps to cover ones back in legal aspects. Often it is recommended to insert a clause into the general terms and conditions allowing filtering for security reasons.
- Making exceptions for customers simplifies the process for the subscribers, but makes it more complicated for the provider. Discussing whether and how exceptions can be made is highly recommended.
- Communicating the change from an opened to a closed port 25 might help especially more experienced users to understand their problems. On the other hand it might scare customers even if they are usually not affected.
- Blocking port 25 at once might seem to be rude and overload the providers' hotlines. An introduction in small parts (e.g. by region) can help to mitigate this problem, but complicates a communication as discussed before.

#### 8.1.1.4 Risks when blocking port 25

Once spammers notice that spamming via bots without using smart hosts has become inefficient, they will change their strategy. A possible and definitely probably solution for them would be to improve the bots letting them send emails via the smarthosts of the providers. As a conclusion of this email servers of providers will get bad reputations, since spam is sent via those. This is a not negligible danger leading to other prevention mechanisms when blocking port 25.

On the other hand, this gives the providers a comprehensive control of the customer's email flow allowing them to identify infected computers easily. They are able to scan the emails (for viruses, spam, etc.) and inform as well as quarantine or even block these customers. Moreover rate limiting, i.e. a limitation of the amount of emails sent by a customer, can help reducing the amount of spam sent via smarthosts.

All in all managing port 25 is a perfect chance to mitigate the botnet problem in a provider's network in order to avoid spam caused by bots.

### 8.1.2 **Consider network level blocking**

Network level blocking mechanisms are as efficient as they are dangerous. Blocking an SMTP connection can highly reduce the consumed technical and human resources. On the other hand, if blocked in a wrong case, i.e. denying a legitimate sender makes it impossible to see the content of the mails which would have been sent via this connection. For this reason especially the usage of blacklists in order to deny connections depending on the quality of the blacklist can be adventurous.

#### 8.1.2.1 Blacklisting

As mentioned earlier, it is very important to consider which blacklists are reliable enough to use them for blocking SMTP connections. A first approach by Al Iverson tried to figure out

---

<sup>86</sup> Publishing should be done on the homepage as well as a contribution to [pbl.spamhaus.org](http://pbl.spamhaus.org).

false positive rates of blacklists<sup>87</sup>, showing the risk of using them. However, only considering bulk email this approach is not very realistic for a user's mailbox. Another approach, considering personalised non-bulk email was discussed in chapter 7.8.2.

Due to these experiences and first findings it can only conditionally be recommended to use blacklists in order to deny entire SMTP connections. A recipient should not use too many IP based lists for blocking in order to keep the false positive rate reasonable low. Moreover, when using blacklisting, it is *strongly* advisable to combine it with a whitelist (as described chapter 5.1.3). The quality of blacklists should be considered well before using them, as described in chapter 7.3.

Blacklists are a very important instrument in the battle against spam. The reputation among mail administrators could not differ more widely, but in the end the majority of almost 90% of all survey participants use blacklisting as it is needed to mitigate the spam problem. Although a blacklist might be too aggressive for rejecting an SMTP connection, it might be useful to be integrated into a scoring system. Such a system usually combines several tests on a message, builds up a score from the single results of all tests and decides then whether an email is likely to be spam. This approach is much softer than blocking the entire SMTP connection. Received emails can be marked as spam and nevertheless be read in case it was a false positive.

#### 8.1.2.2 Whitelisting

Blacklists often contain email servers, which send ham as well as spam messages. It depends on the policies and the knowledge of the administrators if these servers are listed. The receiver should decide whether he wants to block all emails of these servers. If a server is listed on a blacklist used by the recipient and the recipient nevertheless wants to receive email from this server, he has to whitelist this server. This reduces the false positive rate, giving still other filters (like content-filters) the chance to filter out spam. Moreover whitelisting saves the resources needed for blacklisting for well-known senders.

In combination with greylisting, whitelisted senders do not get delayed when sending emails. It is recommended to bypass greylisting for senders listed in well-known whitelists.

#### 8.1.2.3 Greylisting

Greylisting is a very efficient solution in order to block spam and should be used. It has a false positive rate near zero and only spammers and bad configured sending servers will not succeed to bypass greylisting. Since spammers will try to trick greylisting by sending twice (although this is hardly to work, because the embargo time as described in chapter 5.1.2 is usually one hour or more), recipients not using greylisting will suffer even more by receiving duplicate spam.

### 8.1.3 Support sender authentication

As indicated in chapter 5.1.4, sender authentication only constrained helps to reduce spam. In the case the domain's owner set up sender authentication correctly it proves whether the sender is allowed to send from this domain or not. This helps preventing misuse of domains and might identify the sender as the person it claims to be. On the other hand, spammers can and will adopt these techniques for their own use, hoping to have better chances to slip through anti-spam installations. Thus sender authentication will not

---

<sup>87</sup> See <http://stats.dnsbl.com/> for this statistic.

completely avoid spam, but simplifies identifying spammers and protecting domains of misuse.

#### 8.1.3.1 Receiving part – Using information

This part describes possible measures in the field of sender authentication to take when receiving an email, i.e. using given sender authentication data.

For the reasons described before, sender authentication can only conditionally influence the decision whether an incoming email is likely to be spam. A definition by cases helps to tag possible situations and gives advice for appropriate decisions.

- When an email is received and the authentication fails, i.e. the sender is explicitly not allowed to send spam with this address, the email should be refused. This situation requires well-administrated sender authentication data by the claimed sender<sup>88</sup> and helps to mitigate the spam problem by using sender authentication.
- If the claimed sender did not manage authentication records, this should be considered as neutral. In this case the sender can be legitimate or not, since a correct decision or valuation is not possible.
- A successful authentication ensures the sender is authorised to send emails from this domain respectively email address. Because many spammers also use sender authentication<sup>89</sup>, this should not lead to more confidence in an email to be spam. This case only ensures the sender did not misuse a foreign domain. In other words, an authenticated email should neither imply that the email is ham and nor result in better score results in anti-spam systems.

#### 8.1.3.2 Sending part – Providing information

This part describes recommended activities in the field of sender authentication to take before sending emails, i.e. providing sender authentication data.

In the last months the sender authentication scene was very active. Especially the standardisation of DKIM as RFC 4871 at the IETF<sup>90</sup> was a big step towards the future of sender authentication. However, with path-based and signature-based authentication two opposite factions try to establish their standards. It seems to be too late to agree on one standard<sup>91</sup>, so all email servers should support path- as well as signature-based authentication<sup>92</sup>.

The choice between existing path-based authentication methods is difficult, since IETF's MARID working group proposed many standards that are open to discussion. However, it seems to figure out that SPF is a very successful and the most promising candidate for

---

<sup>88</sup> *In other words the person from which the email apparently comes from.*

<sup>89</sup> “[...] that only about 5% of all incoming messages came from domains that published a valid sender authentication record [...]. Within that 5%, slightly more is spam than is legitimate e-mail [...]” and “The idea that SPF would point to legitimate e-mail because spam would fail SPF checks is not true, because spammers have rolled out SPF records, too.” See the full article at <http://www.computerworld.com/softwaretopics/software/groupware/story/0,10801,95617,00.html>.

<sup>90</sup> See <http://www.ietf.org/rfc/rfc4871.txt>

<sup>91</sup> IETF's “MARID” working group tried to develop a DNS-based mechanism for storing and distributing information associated with that authorization, but failed in 2004 to find a consensus. See <http://www.ietf.org/html.charters/OLD/marid-charter.html> for more information.

<sup>92</sup> This situation can be compared with the DVD standards: After several years of trying to find out one standard out of DVD+ and DVD-, usually DVD readers and writers handle both formats.

path-based authentication. Considering the conducted survey, the most used path-based method is SPF, with a usage of almost 50%. Therefore providers should provide SPF records, hoping that this standard will be established as a best practice.

Reviewing signature-based sender authentication methods is much easier. With DKIM the first and as of 2007 only IETF standard for signature-based authentication was created in May 2007. As shown in the results DKIM's ancestor DomainKeys is not widely used yet, but providers plan to implement it. As opposed to path-based sender authentication the effort with signature-based methods is at the sender's side. Therefore especially smaller providers will probably flinch from implementing it and prefer easier implementations like path-based methods. However, if affordable, providers should also implement the new DKIM standard since it mitigates some problems of SPF and has a chance to establish as best practice.

#### **8.1.4 Offer user-defined anti-spam solutions**

An often done mistake by providers is to lump all users together. In other words, the offered anti-spam solutions are equal for each customer, or differ only in specific payment levels as developed in a business model. This is very hazardous, because the most anti-spam methods use a big database of anti-spam signals. These indications often do not fit to customers' needs.

Blacklisting should be combined with user-defined whitelisting. Offering the user to administer a whitelist for his own mailbox makes it possible for him to receive emails from senders even if they are blacklisted. Providing such loopholes the customer can avoid too many false positives especially with very aggressive blacklists.

Content-based filters should also be user defined. A provider can build up a huge database of a scoring system e.g. for phrases. Nevertheless it should be possible for the user to edit respectively enhance these databases. If this is not possible, some fringe groups who deal with spam-related phrases will get discriminated. Although this sounds deceptive, e.g. researches in the field of Viagra or brokers want to receive emails with (for most other users) suspicious content.

Furthermore it is recommendable to allow customers enabling/disabling of the individual anti-spam methods. Giving a user the choice to activate/deactivate the entire anti-spam solution is a first step to this solution. However, a better practice is making the control of individual methods possible. Either enabling users to switch on and off specific methods or setting the level of aggression are very advisable feasibilities.

All of the possibilities mentioned help to optimise the false positive/negative rates for the individual customers' needs. Users will appreciate this level of freedom by a higher acceptance of the anti-spam system, since they can control the level of risk as well as efficiency of the anti-spam solution.

#### **8.1.5 Contribute to anti-spam networks**

As spammers usually work together and exchange new techniques or other kind of important information for their "business", providers should do. Providers not contributing in anti-spam networks harm themselves, because they loose a unique chance to gain more power on fighting spam. There are two types of possible networks, the closed and the open network.

An open network is accessible for everyone, usually without paying something, checking the participant's identity or authorisation process to join. Since usually spammers try to contribute to these networks as well, those have several risks. Nevertheless they are very important and should be used. They allow exchanging non-sensitive information readable by everyone and have the ability to cover therefore a big range of information recipients. On the other hand, neither sensitive data nor data helpful for spammers should be exchanged via an open network. The simplest kinds of such networks are mailing lists, newsgroups or forums available for the publicity.

Closing a network increases the security, so that nobody in the network carries information to recipients not allowed to read it (mostly spammers). They offer a smart way to exchange information through providers and build up a collective knowledge base about spammers, with a very small risk this information will reach frauds. Possible kinds of information include sharing knowledge about anti-spam solutions, distributing data used by individual methods or arranging combined actions to combat spam. On the contrary this exchange excludes the majority of other interested legitimate parties and helps a limited number of participants only.

In spite of that, closed networks between providers have become – as far as obvious – good practice. Very popular closed groups are the private sessions of ISPs' technicians during the MAAWG meetings. The estimated number of unknown cases is high and participating in such groups is highly advisable. Moreover it is very useful to have direct contacts to related providers in order to guarantee fast communications in emergencies. Building up such social networks is up to every email server administrator.

Especially within closed networks a high benefit can be reached without making reasonable effort. Often providers run spamtraps or some kind of frequency analysis in order to catch and examine spam. From received spam many conclusions can be drawn: IP addresses for blacklisting, hash values for checksum comparison or extracting URIs for URIDNSBLs are only a few of them. When sharing them between providers, this data gets multiplied for each provider in a win-win situation. Distributed reputation systems help to adopt this task in an easy manner and research in this field should be done or supported by providers interested in such closed groups.

## **8.2 Controversial anti-spam methods**

### **8.2.1 The future of email data analysis**

Data analysis (like Bayesian filters) plays a huge role in fighting spam. Content-filtering is used by more than 80% of the survey's participants. But the future of this kind of spam filtering is unknown. On the Internet are as many promoters as people with aversion for data analysis. This chapter describes why it is risky to rely on data analysis only.

Content filters work well today only if spammers do not do. A high percentage of a well produced spam wave will pass data analysis and only mistakes made by spammers make it possible to filter it out:

- When disguising their identity spammers forge email envelopes (as described in chapter 5.2.1.1), but do not give much effort to adopt this correctly. If they improved their bulk transmissions, methods scanning for this would fail.
- Image spam makes it nearly impossible to use content filters in an efficient way. First, converting the image into spam consumes many resources. Secondly, after disguising

the image in a very hard manner, it is nearly impossible to convert it to text without running in several mistakes.

- Bayesian poisoning<sup>93</sup> is a good way for spammers to train Bayesian filters in a bad manner. Via inserting ham-like words in spam emails the filter either scores the ham words worse or the spam words better. As a result spam is likely to pass the filters. This sounds easy, but is today still a quite undiscovered topic on the net<sup>94</sup>.

Why should not spammers learn from this in the future and improve their spoofs? Why should not they improve their image spam by adopting up-to-date CAPTCHA mechanisms? What prevents spammers before finding an efficient way how to poison Bayesian filters? These questions cannot be answered by now, but should wake up people using content filters only.

Another until now completely undiscovered branch is the effect of localisation of spam (see chapter 3.2). Until now recipients usually get ham emails in their origin language. On the other side, spam mostly has been sent in English language. This lead to the fact that Bayesian filters usually scored English words worse than words in the origin language. Now the strategy of spammers is changing and they try to adopt the language of their spam depending on their recipients. After doing so, this highly irritates the Bayesian filters. Since the local language usually is well scored in a Bayesian filter, mails in this language are likely to lead to a bad false negatives rate.

Concluding nowadays email data analyses work well, but the future of these is unknown. If used, the email data analysis should only be one part of an anti-spam solution and feed a scoring system that consists of many other checks.

### 8.2.2 Latent damage through SAV

As mentioned in chapter 5.1.5 sender address verification (short: SAV) might lead to big problems regarding DoS attacks. Many places on the Internet discuss the dangers of using SAV in a larger scale. The main problems are spam waves with a unique forged sender's domain, leading to many SAVs (i.e. SMTP dialogs) to this domain. For this some blacklists dedicated for users of SAV are available. SAV has a very bad reputation and should not be used without rigorous restrictions. It tries solving the spam problem with consuming resources of others.

When using SAV, even if it is not recommended, the following additions to the procedure should strongly be considered:

- In any case the results of the SAV should be cached. This prevents for performing the same checks more often than needed.
- Sender authentication methods should be given the preference before using SAV. Authentication methods use DNS instead of SMTP and are therefore much lighter than SAV. A negative authentication will lead to a rejection and a positive authentication should consider the email address as correct. Only a neutral authentication, i.e. a case where no decision was possible, might get checked with SAV.

---

<sup>93</sup> As described before in chapter 5.2.2. For details see

[http://en.wikipedia.org/wiki/Bayesian\\_poisoning](http://en.wikipedia.org/wiki/Bayesian_poisoning)

<sup>94</sup> Some approaches of Bayesian poisoning are discussed on the web, showing that until now there is little effect of poisoning: See <http://www.cs.dal.ca/research/techreports/2004/CS-2004-06.shtml> and <http://www.virusbtn.com/spambulletin/archive/2006/02/sb200602-poison> for details.

- SAV using servers must not be listed on any blacklists<sup>95</sup> and provide information required by the SAV tested server, e.g. a reverse DNS record.

The given limitations will limit the harms when using SAV. Moreover it encourages domain owners to administrate their own valid authentication data as described in chapter 8.1.3.2 in order to keep SAV traffic low. However, the usage of SAV is generally discredited and cannot be recommended. Performing such heavy tests will bring providers into disrepute. It is a debatable point whether spammers will simply improve their spam with valid sender email addresses undermining the benefit of SAV.

### 8.2.3 (R)evolution of structural adjustments

Each type of structural adjustment provides a special attraction. Usually, if applied on every part on the Internet, these adjustments solve all spam problems as a kind of panacea. The bitter truth is that medium- or even long-termed none of these methods can and will be widely implemented.

Most of the adjustments described in chapter 5.3 require large modifications of the structure. But history has shown that revolutions are very hard and unlikely to achieve. It is impossible to change the current architecture in a way making it incompatible to the old one. That leads to the fact, that if structural adjustments will have success, this will happen more slowly. Such evolutions were very successful in email's history, demonstrated with examples like the SMTP service extensions<sup>96</sup>.

An e-postage approach requires a micro-payment system (as described in chapter 5.3.4) and will hardly be accepted by users. Those got to know email as a cost-free system. Probably their psychology will not accept any changes making this system partly monetary. Moreover creating and managing a fast, ubiquitous and secure micro-payment system is a challenging and until now unsolved problem.

Challenge response mechanisms are also hardly acceptable for email senders. They are used to mail a message without doing any further steps, expecting the recipient to get the email. Changing this behaviour for email would introduce a further process making electronic communications more asynchronous as it is nowadays. Moreover the user has to make more efforts to send an email. Therefore challenge response mechanisms are not a solution for the worldwide spam problem. However, they could be very useful for entities who suffer hardly from receiving spam and who are in a position to expect legitimate users bringing up this overhead.

Proof-of-work mechanisms (as described in chapter 5.3.2) have an unsolved problem with systems sending legitimate bulk emails, e.g. mailing lists or newsletters. For instance Hashcash widely discusses this problem in the official FAQ<sup>97</sup>, claiming they found a solution. There are general two approaches to solve this issue, but both of them have important disadvantages. The first idea, letting bulk senders proof some work for each email, is not affordable. And if it was, spammers could use this method as well. A second approach aims at whitelisting the mailing list at the client side. However, once a proof-of-work

---

<sup>95</sup> Especially specialised blacklists like <http://www.backscatterer.org/> listing servers using SAV must not list the server using SAV, which might be a contradiction.

<sup>96</sup> ESMTP, as described in RFC 1869, "does not require modification of existing SMTP clients or servers unless the features of the service extensions are to be requested or provided".

<sup>97</sup> See <http://www.hashcash.org/faq/> in chapter 5 d)-f) for more details.



mechanism with such a configuration was established, spammers would use this security hole by sending via mailing lists. Due to these concerns proof-of-work mechanisms are in general a nice idea to improve the reliability of some senders by adding an attest for his will to send an email, but it will not be a panacea for the entire spam problem.

To summarise, all structural adjustments promise to be a solution for solving any spam issues. On the other hand, they run either into danger to be not accepted by the users or are weak at some points allowing spammers to slip through. Instead of longing for revolutions eagerly, anti-spammers should try to apply compatible and well-considered structural adjustments. These include managing port 25 (as recommended in chapter 8.1.1) as well as traffic shaping (as described in chapter 5.3.5). Other, coarser changes have a highly doubtful future.

## **9 Conclusion**

Although spam exists since many years, there is no optimal solution for the problem. Spammers improve their methods and bypass outdated anti-spam installations. Recent trends jeopardise multiple measures currently applied. Domain tasting is a big danger for sender authentication, since spammers can and will use authenticated emails in future. Image spam and the localisation of the emails may compromise content data analyses, like the famous Bayesian filter. These challenges will play a role in the mid-term future of spam and require every email server administrator to be updated.

This document gives pros and cons of each current anti-spam method and may give some new ideas for providers. Administrators should verify whether they use or are willing to use the not-recommended technologies. Moreover, the alarming trends should make provider thinking of their current anti-spam installations as a matter of prudence. Since a rolling stone gathers no moss, they should proactively watch out for future incidents and have forward-looking anti-spam setups.

The big amount of spam coming from botnets shows that most providers care rather about incoming than outgoing spam. Convincing the management as well as the marketing department of spending costs for anti-spam solutions that prevent outgoing spam only is very hard work. However, the technical opportunities, first of all managing port 25, are available since many years and lack of publicity only. Botnet operators are getting international and Europe is going to lose a good reputation for its (former) low spam ratios. In order to circumvent a bad reputation, only efficient steps against botnets will help.

Following the conducted survey about a half of participants already managed to block port 25, which eliminates spam sent via botnets. This shows still big potential and encourages other providers to do the same job. The providers declared spam as the second highest internet security threat, emphasising the needs on some actions. Additionally the survey showed only small popularity of sender authentication, which might be increased by the standard of DKIM and rising usage of SPF. Last but not least the survey gave a good impression about best practices on anti-spam. It stresses blacklisting as the most used and very efficient method against illegitimate emails.

Detailed research on this topic has been done in order to enlighten current blacklists. It figured out that quality assurance is a very important topic for blacklists. Many mail server administrators wonder which blacklist to trust and use. This paper describes methods how to choose blacklists with a minimal risk. It discusses possibilities for blacklist/whitelist operators in order to allow qualified data. The intersections between blacklists showed relationships between multiple lists. The weak coverage by blacklists with less than 20% of the advertised IPv4 address space can still be improved.

Further research will be done on the quality of blacklists, especially by implementing the draft of a hamtrap and doing research on most-efficient ratios between true positive and low false positive rates. Moreover it is planned to investigate and publish the presented data online on a regular basis. Giving this kind of information to providers will allow them to compare between blacklists and choose them for an optimal usage.

Summarising the combat against spam has not been won yet. Spammers are usually one step ahead the average providers and circumvent anti-spam installations. However, this

paper worked out best practices for providers and analysed the future trends of spam for proactive arrangements. The given recommendations based on determined best practices of the providers will help to mitigate the spam problem. Moreover the detailed research on the very famous blacklisting is a help for providers to improve the efficiency and coincidentally minimise the risk of blacklists.

Eventually spam is like rain - you cannot stop it, but you can take an umbrella in order to stay dry and enjoy your walk.

## A Glossary

ASN	An <b>A</b> utonomous <b>S</b> ystem <b>N</b> umber identifies an Autonomous System uniquely.
ASRG	The <b>A</b> nti- <b>S</b> пам <b>R</b> esearch <b>G</b> roup is part of the IRTF and its work areas include new or improved anti-spam tools and techniques, administrative tools and techniques, evaluation frameworks and measurement, and approaches that involve changes to the existing applications and protocols.
Backscattering	Spammers often use forged email addresses to send emails. Once the spam email has been accepted, but cannot be stored into a users account, a bounce messages is created to inform the sender about the failure. Receiving many of such bounce messages (backscatter) is called backscattering.
Bayesian filter	Bayesian spam filtering is the process of using a Naïve Bayes classifier to identify spam email. Since then it has become a popular mechanism to distinguish illegitimate spam email from legitimate email (sometimes called ham). Bayes's theorem, in the context of spam, says that the probability that an email is spam, given that it has certain words in it, is proportional to the probability of finding those certain words in spam email.
BCP	A paper called <b>B</b> est <b>C</b> urrent <b>P</b> ractices gives suggestions how to apply special methods in a generally most logical choice.
Bulk email(s)	Bulk emails are large volumes of emails. Bulk email is only considered to spam, when it's not unsolicited. See definition in chapter 2.1.1.
Bounce	A bounce message is a report of the receiving MTA sent to a former sender, when the receiving MTA accepted an email from this sender and is not able to deliver it to the stated recipient.
CAPTCHA	A <b>C</b> ompletely <b>A</b> utomated <b>P</b> ublic <b>T</b> uring test to tell <b>C</b> omputers and <b>H</b> umans <b>A</b> part is a type of challenge-response test used in computing to determine whether or not the user is human. Most common application of CAPTCHA is disguising text into an image that is readable by humans only.
Checksum	A checksum is a simple way to protect the integrity of data by de-

tecting errors in data. The same operation on the data (checksum function) will lead to the same result, if the data has not been modified or otherwise different.

- DCC** The **D**istributed **C**hecksum **C**learinghouse centrally collects checksums of spam emails to allow other parties requesting this database for ingoing emails. If an email hits an existing checksum, it can be more easily identified as spam.  
Website: <http://www.rhyolite.com/anti-spam/dcc/>
- DNS** The **D**omain **N**ame **S**ystem allows associating several sorts of information with a domain name, e.g. IP addresses or textual information.
- DNSBL** A **DNS** **B**lacklist is a list of IP addresses and can be easily queried by computer programs on the Internet, to avoid communication with unwanted parties. Spam sending servers/clients are likely to get listed on a DNSBL.
- DNSWL** A **DNS** **W**hitelist is a list of IP addresses and can be easily queried by computer programs on the Internet, to reduce the chances of false positives while spam filtering. Ham sending servers are likely to get listed on a DNSWL.
- DoS** A **D**enial-**o**f-**S**ervice attack is an attempt to make a computer resource unavailable to its intended use.
- ENISA** The **E**uropean **N**etwork and **I**nformation **S**ecurity **A**gency is an agency of the EU and has its seat in Heraklion, Crete (Greece). Objective of ENISA is to improve network and information security in the European Union.  
  
Website: <http://www.enisa.europa.eu/>
- ESMTP** **E**xtended **S**MTP is a definition of protocol extensions to the SMTP standard. These service extensions are support of TLS, authentication and command pipelining and others. The list of new parameters is at <http://www.iana.org/assignments/mail-parameters>, the RFC can be found at <http://tools.ietf.org/html/rfc1869>.
- ESP** Email Service Provider, see chapter 2.2.1 for details.
- EU** The **E**uropean **U**nion is with 27 member states one of the largest

economic and political entities in the world. It is the successor of the European Economic Community and was established in 1992.

Fuzzy Checksum	See “local sensitive hash functions”
False negative	Spam email that was wrongly not marked or filtered as spam.
False positive	Ham email that was marked or filtered as spam by mistake.
Hash function	A hash function is a reproducible method of turning some kind of data into a (relatively) small number that may serve as a digital fingerprint of the data. It can be used for checksum, to check the data integrity.
HTML	<b>H</b> ypertext <b>M</b> arkup <b>L</b> anguage is the common markup language for the creation of web pages. Since emails mostly are not plain text, they contain HTML code to display different fonts, colors, images, etc.
ICANN	The <b>I</b> nternet <b>C</b> orporation for <b>A</b> ssigned <b>N</b> ames and <b>N</b> umbers is a non-profit organisation in order to oversee a number of Internet-related tasks. The tasks of ICANN include managing the assignment of domain names and IP addresses. Website: <a href="http://www.icann.org/">http://www.icann.org/</a>
IRTF	The <b>I</b> nternet <b>R</b> esearch <b>T</b> ask <b>F</b> orce has the mission to “promote research of importance to the evolution of the future Internet by creating focused, long-term and small Research Groups working on topics related to Internet protocols, applications, architecture and technology”.
ISP	<b>I</b> nternet <b>S</b> ervice <b>P</b> rovider, see chapter 2.2.1 for details.
LMTP	The <b>L</b> ocal <b>M</b> ail <b>T</b> ransfer <b>P</b> rotocol, as described in RFC 2033, is an alternative to ESMTP for limited circumstances in which it is desirable to implement a system where a mail receiver doesn’t manage a queue.
Local sensitive hash functions	A local sensitive hash functions (sometimes called “fuzzy checksum”) is a checksum which tolerates little modification on the data by giving the same result as before the modification. It can be

used for checksum based anti-spam methods.

MAAWG	The <b>M</b> essaging <b>A</b> nti- <b>A</b> buse <b>W</b> orking <b>G</b> roup is a global organisation focusing on fighting spam, phishing and other possible forms of email abuse. It has a broad base of ISPs and network operators representing over 600 million mailboxes. Website: <a href="http://www.maawg.org/">http://www.maawg.org/</a>
MARID	The <b>M</b> TA <b>A</b> uthorization <b>R</b> ecords <b>I</b> n <b>D</b> NS working group was team established by IETF tasked to propose standards for e-mail authentication. SPF, Reverse MX and MTAMARK are the most well-known proposals made by MARID.
MDA	<b>M</b> ail <b>D</b> elivery <b>A</b> gent, see chapter 2.2.4 for details.
MSA	<b>M</b> ail <b>S</b> ubmission <b>A</b> gent, see chapter 2.2.4 for details.
MTA	<b>M</b> ail <b>T</b> ransfer <b>A</b> gent, see chapter 2.2.4 for details.
MUA	<b>M</b> ail <b>U</b> ser <b>A</b> gent, see chapter 2.2.4 for details.
PDF	<b>P</b> ortable <b>D</b> ocument <b>F</b> ormat is an open standard for a file format created by Adobe Systems for worldwide readability.
OCR	<b>O</b> ptical <b>C</b> haracter <b>R</b> ecognition is a type of computer software designed to translate images of handwritten or typewritten text into machine-editable text. In the region of spam it's used to convert image spam into text.
Reverse DNS	Reverse DNS, often called rDNS, makes it possible to determine via DNS hostnames associated with a given IP address.
RFC	The Internet Engineering Task Force (IETF) usually uses <b>R</b> quest for <b>C</b> omments for proposing/publishing new standards. RFCs are documents with a specific, text-like format.
Smart host	A smart host is a type of mail relay server which allows an SMTP server to route emails to an intermediate mail server instead of directly to the recipient's server.
SMTP	<b>S</b> imple <b>M</b> ail <b>T</b> ransfer <b>P</b> rotocol is the standard for email transmis-

sions across the Internet. It is a text-based protocol, where one or more recipients of a message are specified and then the message text is transferred.

Website of definition: <http://tools.ietf.org/html/rfc2821>

Spamtrap	Spamtraps are usually email addresses that are created not for communication, but rather to lure spam. Since no email is solicited by the owner of this spamtrap email address, any email messages sent to this address are immediately considered unsolicited.
Tarpit	Service, usually on a server, delaying incoming connections for as long as possible in order to slow down the service for attackers.
TLD	The <b>T</b> op <b>L</b> evel <b>D</b> omain describes the last part of an Internet domain name. Most common TLDs are for instance .com, .org and .net.
Unsolicited email(s)	Unsolicited emails had not been requested by the recipient and can so, in the field of bulk email, be considered as spam. For more details see chapter 2.1.1.
URI	A <b>U</b> niform <b>R</b> esource <b>I</b> dentifier in the field of anti-spam is similar to an URL representing a web address.
URIDNSBL	<b>U</b> RI <b>D</b> NS <b>B</b> lacklist, see chapter 5.2.3 for details.



## B References

As far as references were not mentioned directly this list gives an overview of abbreviated documents:

Reference	Document
[ENISA1]	First ENISA survey in February 2006: <a href="http://enisa.europa.eu/doc/pdf/deliverables/enisa_security_spam.pdf">http://enisa.europa.eu/doc/pdf/deliverables/enisa_security_spam.pdf</a>
[ENISA2]	Second ENISA survey in June 2006: <a href="http://enisa.europa.eu/doc/pdf/deliverables/enisa_security_spam_part2.pdf">http://enisa.europa.eu/doc/pdf/deliverables/enisa_security_spam_part2.pdf</a>
[Comm01]	<a href="http://www.commtouch.com/documents/Commtouch_2006_Spam_Trends_Year_of_the_Zombies.pdf">http://www.commtouch.com/documents/Commtouch_2006_Spam_Trends_Year_of_the_Zombies.pdf</a>
[EStat01]	<a href="http://ec.europa.eu/public_opinion/archives/ebs/ebs_249_en.pdf">http://ec.europa.eu/public_opinion/archives/ebs/ebs_249_en.pdf</a>

## C Annex - ENISA survey

### Security

1. Please **order** the following list of **threats** from most concerning (1) to least concerning (8) for your organisation:
- Act of nature beyond control
  - BGP Hijack\*
  - DNS attacks
  - DoS\*/DDoS\*
  - Social Engineering/Spying
  - Spam
  - Viruses
  - Worms
2. Which of the following **organizational** measures do you take to secure your services?
- We provide regular information on information security to our subscribers by ...
- ... publishing information on our web site
  - ... sending physical mail
  - ... sending email
- We provide detailed written guidance for
- ... staff
  - ... partners
  - ... subscribers  ... including a policy defining permitted/prohibited uses of the messaging services
- We provide security software for users
- ... free of charge
  - ... for a fee
- We provide clear contact details
- ... for email abuse
  - ... for security violations
- We provide security support via Hotline/Helpdesk
  - We provide remote technical assistance (i.e. with access to the device)
  - We maintain up-to-date reverse DNS records
  - We provide training or awareness campaigns
  - We inform subscribers about the legal consequences of sending spam
  - We forbid spamming in Terms & Conditions
  - Other (please specify):
3. Which of the following **technical** measures do you take to secure your services?
- Basic filtering (e.g. spoofed IP address) in ...
- ... Ingress\*
  - ... Egress\*
- Content filtering\* (e.g. anti-viruses) in ...
- ... Ingress\*
  - ... Egress\*
- Quarantining an infected / malicious PC\*
  - Blackholing/Sinkholing\*
  - Other (please specify):
  - Traffic Shaping / Throttling\*
  - DNSSEC\* (RFC 4033-4035)
4. What measures do you take to **become aware** of security or spam problems?
- We track complaints
  - We monitor for traffic peaks

## Security

We deploy real-time traffic anomaly\* and/or signature-based detection
  We subscribe to security intelligence services

We deploy spamtraps\*
  Others (please specify)

5. How do you ensure an **appropriate level** of security? Choose your three most appropriate options:

We follow guidance in international standards
  We follow guidance in national legislation

We follow industry best practice
  We follow the advice of the national computer security organization of our country

We define a Risk Management process
  We define an appropriate level in our Security Policy

We define a Service Level Agreement (SLA)
  We do what is necessary based on our internal risk assessments

No guidance or any measures needed

6. When a particular **risk of a security breach** arises in your network, what do you do?

We inform subscribers directly (i.e. individually)

We inform subscribers via a customer portal (to which only subscribers have access)

We issue reports to the public (e.g. with a press release, on our public website)

We report to the NRA\*

We decide whether and how subscribers should be informed

7. If the risk lies **outside the scope** of the measures you can take directly, then what do you do?

We inform subscribers of any possible remedies that *they* can take

We inform subscribers of any possible remedies that *they* can take and the costs

We inform subscribers of the risk of not implementing counter measures

We stop servicing non-compliant subscribers

8. Regarding protection of **network integrity**, what do you do?

We have a Business Contingency\* (BC) process

We have a Disaster Recovery\* (DR) process

We have a Risk management process

We regularly conduct BC/DR tests (at least yearly)

We do nothing, but we wish we could do more

Others (please specify)

## SPAM

9. Which of the following measures do you take to prevent your subscribers from **sending** spam?

We use a blocking list system:

We put a subscriber on a blacklist\* if the subscriber repeatedly sends spam

We put on a whitelist\* all subscribers who do not send spam

We use a greylist\* system

We block access to port 25 from all hosts on our network other than those that are explicitly authorized to perform SMTP relay functions

We provide Email Submission services on port 587 (as described in RFC 4409)

We limit high outbound mail volumes

<b>SPAM</b>
-------------

We perform outbound virus scanning  
 Other (please specify)

10. Which of the following measures do you take to protect your subscribers from **receiving** unsolicited communications (spam)?

We offer spam-filtering on our network (e.g. by subscribing to black/white-lists) ...

... free-of-charge                       ... for an additional fee

We offer spam-filtering software that subscribers can install on their computers ...

... free-of-charge                       ... for an additional fee

We do nothing, but we wish we could do more  
 Other (please specify)

11. Which of the following **spam-filtering measures** do you take on your network?

<input type="checkbox"/> Blacklisting* (e.g. DNSBL) <input type="checkbox"/> Whitelisting* (e.g. DNSWL or CSA*) <input type="checkbox"/> Greylisting* <input type="checkbox"/> Sender authentication* <input type="checkbox"/> Checksum analysis (e.g. DCC*) <input type="checkbox"/> Blacklisting of URIs (e.g. URIDNSBL)	<input type="checkbox"/> Content filtering* (e.g. rule-based or statistical) <input type="checkbox"/> Reputation system* <input type="checkbox"/> Slowing down the senders connection <input type="checkbox"/> Frequency analysis of connection problems <input type="checkbox"/> Outsourced system (technology unknown) <input type="checkbox"/> Other (please specify)
---	---

12. Which of the following **sender authentication mechanisms** do you implement?

<input type="checkbox"/> SMTP AUTH (RFC 2554) <input type="checkbox"/> POP3 before SMTP <input type="checkbox"/> Sender ID Framework (SIDF*) <input type="checkbox"/> DomainKeys Identified Mail (DKIM*)	<input type="checkbox"/> SMTP TLS (RFC 3207) <input type="checkbox"/> Reverse-MX <input type="checkbox"/> Sender Policy Framework (SPF*) <input type="checkbox"/> Other (please specify)
---	---

13. When do you **analyze** where spam comes from?

On request from ISPs who received spam from our network  
 When an automatically monitored spam level reaches a certain threshold  
 Following complaints from our subscribers  
 We do not analyze where spam comes from  
 Other (please specify)

14. What sort of measures do you take if you detect **spam coming from another ISP**?

We contact the ISP from which this spam originates to discuss countermeasures  
 We filter or block SMTP traffic from that ISP immediately  
 We filter or block SMTP traffic from that ISP if the ISP itself does not take any measures  
 We filter or block IP addresses from that ISP immediately  
 We filter or block IP addresses from that ISP if the ISP itself does not take measures  
 We inform the other ISPs National Regulatory Authority  
 We inform our National Regulatory Authority  
 We pursue legal actions  
 There is not much that we can do  
 Other (please specify)

15. Could you provide us the following information about your anti-spam system?

a) % of aborted connections due to <b>blacklisting</b> (in relation to all SMTP connections):	%
b) % of aborted connections due to <b>unknown recipients</b> (in relation to all SMTP connections):	%
c) % of aborted connections due to <b>greylisting</b> (in relation to all SMTP connections):	%
d) % of accepted connections due to <b>whitelisting</b> (in relation to all SMTP connections):	%
e) % of as virus infected filtered emails (in relation to accepted emails):	%

**SPAM**

f) % of as spam detected filtered emails (in relation to accepted emails): %

16. Do you think that there is a conflict between ISP obligations of delivering messages/protection of privacy and the use of spam filters that block some messages?  
 No  Yes, please specify:
17. Do you plan to install or implement an anti-spam method in the next six months?  
 No  Yes, please specify:
18. How do you process abuse reports?  
 They are processed manually  
 We use the ARF\* standard reporting format  
 We provide feedback loops to other organisations  
 We use another reporting format or automated tools/method, please specify:

**Miscellaneous**

19. If one or several questions did not offer appropriate answer options, please use this space to explain. Please also indicate the number of the question.

20. Do you think that a workshop on the matter of this study before the end of 2007 would be valuable for you?  yes  no

If yes, what would be areas that should be covered:

- Discuss measures of providers regarding spam measures  
 Discuss measures of providers regarding security measures  
 Laws and legacy problems regarding spam  
 Laws and legacy problems regarding security  
 Presentation of study results  
 Presentation from security and spam filtering vendors  
 Presentation of new methods from security and spam filtering research  
 Others (please specify)

21. Could you provide us with the following information that will increase the quality of the survey results?

How many email boxes do you manage?

How many email messages do you transport per day?

Could you indicate the percentage of your helpdesk calls that concern spam? %

How many persons are fully dedicated security staffs?

What is your annual budget in the area of security? EUR

## D Annex – IP address based blacklist entries by country

The following pages show the blacklist entries assigned to countries. Each blacklist is on a separate page, listing the top 50 countries ordered by the sum of listed net ranges.

IP based DNSBLs that have been reviewed are:

- all.dnsbl.sorbs.net
- UCEPROTECT - Level 1
- NiX Spam
- dnsbl.ahbl.org
- sbl.spamhaus.org
- dnsbl.njabl.org
- CBL
- pbl.spamhaus.org
- xbl.spamhaus.org
- dnswl.org
- union of all blacklists

For the origin of the data please see chapter 7.2. If the assignment between range and country wasn't possible, the country column is empty.

The tables list amounts by country. *Rank* indicated the position of the country ordered by the sum of listed net ranges. *Entries* indicates the number of net range entries in the table. Contrarily *Range* sums the size of all net ranges and gives information about how many particular hosts (single IP addresses) have been listed. *Quota* gives information about the percentage of the covered range of a country, i.e. which ratio of the amount of a country's assigned IP addresses is listed in a blacklist.

Rows that have a grey background are countries within the European Union.

**all.dnsbl.sorbs.net**

rank	country	entries	range	quota
1	United States	290528	1216697	0.09%
2	(unknown)	2221	534412	n/a
3	China	458286	523821	0.43%
4	Korea, Republic of (South)	277088	277088	0.50%
5	Canada	31046	228929	0.31%
6	European Union (can apply to any country in Europe)	19966	216571	0.18%
7	India	147009	147009	1.58%
8	Brazil	143033	143033	0.67%
9	South Africa	4074	135399	1.25%
10	Germany	129619	129619	0.19%
11	Turkey	128145	128145	1.55%
12	Spain	118417	118417	0.59%
13	Japan	38929	104464	0.07%
14	Australia	16218	84307	0.26%
15	Italy	77112	77112	0.32%
16	France	70677	70677	0.11%
17	Singapore	4472	70007	1.87%
18	United Kingdom	64815	64815	0.08%
19	Argentina	63792	63792	1.39%
20	Mexico	46795	50890	0.31%
21	Russian Federation	45409	45409	0.29%
22	Poland	44840	44840	0.37%
23	Romania	42519	43794	0.81%
24	Viet Nam	40593	40593	1.17%
25	Chile	31649	31649	0.80%
26	Taiwan (, Province Of China)	29343	29343	0.16%
27	Philippines	29180	29180	1.26%
28	Israel	28819	28819	0.71%
29	Egypt	25375	25375	1.21%
30	Morocco	22973	22973	3.92%
31	Peru	19545	19545	1.91%
32	Portugal	19240	19240	0.56%
33	Thailand	18779	18779	0.52%
34	Malaysia	18381	18381	0.59%
35	Colombia	16966	16966	0.61%
36	Hungary	14020	14020	0.44%
37	Sweden	12549	12549	0.08%
38	Pakistan	12381	12381	2.05%
39	Greece	12146	12146	0.48%
40	Hong Kong	11421	11421	0.17%
41	Saudi Arabia	10972	10972	0.73%
42	Netherlands	10247	10247	0.05%
43	Macedonia (The Former Yugoslav Republic of)	8538	8538	4.34%
44	Ukraine	8429	8429	0.35%
45	Bulgaria	8366	8366	0.35%
46	Switzerland	8292	8292	0.12%
47	Indonesia	7180	7180	0.33%
48	Czechoslovakia (former)	6681	6681	1.01%
49	Lithuania	6570	6570	0.34%
50	Iran, Islamic Republic of	6441	6441	0.48%

**UCEPROTECT - Level 1**

rank	country	entries	range	quota
1	United States	86824	86824	0,01%
2	Poland	56237	56237	0,47%
3	China	55118	55118	0,05%
4	Brazil	49877	49877	0,23%
5	Korea, Republic of (South)	46188	46188	0,08%
6	Germany	43721	43721	0,06%
7	India	34301	34301	0,37%
8	France	32724	32724	0,05%
9	Turkey	31316	31316	0,38%
10	Russian Federation	31137	31137	0,20%
11	Taiwan (, Province Of China)	21466	21466	0,12%
12	Thailand	19441	19441	0,54%
13	United Kingdom	15567	15567	0,02%
14	Mexico	15303	15303	0,09%
15	Egypt	15104	15104	0,72%
16	Israel	14918	14918	0,37%
17	Spain	13220	13220	0,07%
18	Morocco	12700	12700	2,17%
19	Argentina	12425	12425	0,27%
20	Viet Nam	11792	11792	0,34%
21	Chile	11015	11015	0,28%
22	Italy	10301	10301	0,04%
23	Romania	10228	10228	0,19%
24	Peru	9175	9175	0,90%
25	Philippines	8624	8624	0,37%
26	Colombia	7569	7569	0,27%
27	Malaysia	7344	7344	0,24%
28	Hungary	7277	7277	0,23%
29	Ukraine	6469	6469	0,27%
30	Japan	5702	5702	0,00%
31	Portugal	4815	4815	0,14%
32	Canada	4621	4621	0,01%
33	Netherlands	4332	4332	0,02%
34	European Union (can apply to any country in Europe)	3418	3418	0,00%
35	Czech Republic	3375	3375	0,07%
36	Australia	3239	3239	0,01%
37	South Africa	3226	3226	0,03%
38	Bulgaria	3209	3209	0,14%
39	Greece	3159	3159	0,12%
40	Switzerland	2912	2912	0,04%
41	Hong Kong	2882	2882	0,04%
42	Singapore	2650	2650	0,07%
43	Slovakia (Slovak Republic)	2617	2617	0,19%
44	Indonesia	2507	2507	0,11%
45	Sweden	2369	2369	0,01%
46	Austria	2266	2266	0,03%
47	Croatia (Hrvatska)	2209	2209	0,25%
48	Dominican Republic	2132	2132	0,91%
49	Algeria	1981	1981	0,73%
50	Pakistan	1956	1956	0,32%



**NiX Spam**

rank	country	entries	range	quota
1	United States	12278	12278	0,00%
2	Germany	5775	5775	0,01%
3	Korea, Republic of (South)	5746	5746	0,01%
4	China	4650	4650	0,00%
5	Russian Federation	4401	4401	0,03%
6	Brazil	3444	3444	0,02%
7	Turkey	3141	3141	0,04%
8	France	3122	3122	0,01%
9	Poland	2799	2799	0,02%
10	United Kingdom	2185	2185	0,00%
11	Mexico	1878	1878	0,01%
12	Spain	1856	1856	0,01%
13	India	1782	1782	0,02%
14	Israel	1434	1434	0,04%
15	Argentina	1301	1301	0,03%
16	Italy	1156	1156	0,01%
17	Romania	1130	1130	0,02%
18	Thailand	1129	1129	0,03%
19	Japan	972	972	0,00%
20	Peru	889	889	0,09%
21	Chile	869	869	0,02%
22	Taiwan (, Province Of China	864	864	0,01%
23	Netherlands	754	754	0,00%
24	Viet Nam	738	738	0,02%
25	Canada	703	703	0,00%
26	Morocco	673	673	0,12%
27	Colombia	672	672	0,02%
28	Ukraine	662	662	0,03%
29	Philippines	632	632	0,03%
30	Hungary	626	626	0,02%
31	Egypt	591	591	0,03%
32	Czech Republic	567	567	0,01%
33	Bulgaria	489	489	0,02%
34	Portugal	424	424	0,01%
35	Saudi Arabia	419	419	0,03%
36	European Union (can apply to any country in Europe)	413	413	0,00%
37	Malaysia	409	409	0,01%
38	Switzerland	335	335	0,01%
39	Sweden	333	333	0,00%
40	Greece	287	287	0,01%
41	Australia	264	264	0,00%
42	Hong Kong	258	258	0,00%
43	Austria	253	253	0,00%
44	Venezuela	238	238	0,01%
45	Dominican Republic	213	213	0,09%
46	Denmark	205	205	0,00%
47	Slovakia (Slovak Republic)	202	202	0,02%
48	Lithuania	200	200	0,01%
49	Indonesia	199	199	0,01%
50	Latvia	194	194	0,02%

**dnsbl.ahbl.org**

rank	country	entries	range	quota
1	China	835075	1134078	0.94%
2	United States	378388	413344	0.03%
3	Korea, Republic of (South)	327691	327946	0.59%
4	Brazil	172302	172302	0.81%
5	India	107812	107812	1.16%
6	France	100731	100731	0.16%
7	Argentina	95347	95347	2.08%
8	Mexico	93367	93367	0.57%
9	Taiwan (, Province Of China)	80235	80235	0.43%
10	Spain	58504	75142	0.38%
11	Japan	74636	74636	0.05%
12	Germany	73414	73414	0.11%
13	Chile	50114	50114	1.26%
14	Canada	49238	50003	0.07%
15	Turkey	43816	44071	0.53%
16	United Kingdom	42416	42671	0.05%
17	Malaysia	38114	38114	1.22%
18	Italy	35208	35718	0.15%
19	Israel	31641	31641	0.78%
20	Poland	27173	27173	0.23%
21	Peru	24676	24676	2.41%
22	Thailand	22034	22034	0.61%
23	Singapore	21722	21722	0.58%
24	Portugal	21229	21229	0.62%
25	Hong Kong	20623	20623	0.30%
26	Belgium	20088	20088	0.37%
27	Russian Federation	18487	18487	0.12%
28	Australia	18262	18262	0.06%
29	Colombia	18169	18169	0.65%
30	European Union (can apply to any country in Europe)	13754	13754	0.01%
31	Austria	13651	13651	0.20%
32	Netherlands	13405	13405	0.07%
33	Sweden	11042	11297	0.07%
34	Philippines	9123	11170	0.48%
35	Switzerland	10448	10448	0.15%
36	Hungary	9841	9841	0.31%
37	Romania	8436	8436	0.16%
38	Pakistan	8148	8148	1.35%
39	(unknown)	7817	7817	n/a
40	Morocco	6870	6870	1.17%
41	Denmark	6269	6269	0.08%
42	Venezuela	6230	6230	0.18%
43	South Africa	5426	5489	0.05%
44	Greece	4697	4697	0.18%
45	Indonesia	4604	4604	0.21%
46	Slovenia	4535	4535	0.43%
47	Czech Republic	4265	4265	0.09%
48	Norway	4139	4139	0.06%
49	Viet Nam	3916	3916	0.11%
50	Egypt	3617	3617	0.17%

**sbl.spamhaus.org**

rank	country	entries	range	quota
1	United States	1592	627451	0.05%
2	China	326	281022	0.23%
3	(unknown)	69	241147	n/a
4	Russian Federation	236	181160	1.36%
5	Australia	59	134779	0.41%
6	European Union (can apply to any country in Europe)	87	73914	0.06%
7	Romania	72	43854	0.40%
8	India	85	36997	0.04%
9	Canada	142	23491	0.10%
10	Taiwan (, Province Of China	111	19156	0.03%
11	Korea, Republic of (South)	169	16418	0.06%
12	Netherlands	135	11047	0.20%
13	Argentina	106	8995	25.81%
14	Belize	6	8198	0.20%
15	Thailand	76	7087	0.00%
16	Japan	162	6227	0.12%
17	Czech Republic	44	5503	0.03%
18	Ghana	7	5122	4.11%
19	Brazil	84	5015	0.13%
20	Viet Nam	20	4625	0.01%
21	United Kingdom	172	4300	0.08%
22	Colombia	26	4151	0.15%
23	Venezuela	3	4098	0.10%
24	Belarus	2	4097	0.06%
25	Israel	62	3883	2.59%
26	Hong Kong	84	3877	0.12%
27	Nigeria	29	3370	8.45%
28	Ukraine	29	3366	0.16%
29	Spain	65	3260	1.46%
30	France	109	3183	0.14%
31	Senegal	15	3142	0.02%
32	Philippines	34	2881	0.01%
33	Denmark	25	2341	0.03%
34	Mexico	48	1877	0.01%
35	Lithuania	18	1515	0.08%
36	Austria	16	1292	0.00%
37	Germany	140	1056	0.01%
38	South Africa	16	1043	0.53%
39	Lebanon	4	1027	100.20%
40	Virgin Islands (British)	1	1024	0.01%
41	Turkey	40	984	1.38%
42	Poland	76	866	0.01%
43	Burkina Faso	12	780	5.97%
44	Cote D'Ivoire (Ivory Coast)	14	779	0.02%
45	New Zealand (Aotearoa)	10	775	0.04%
46	Egypt	30	767	0.00%
47	Italy	85	757	0.05%
48	Iran, Islamic Republic of	25	660	0.02%
49	Malaysia	11	553	0.00%
50	Sweden	30	547	0.05%

**dnsbl.njabl.org**

rank	country	entries	range	quota
1	China	1263431	1263431	1.05%
2	Korea, Republic of (South)	447158	447158	0.81%
3	Brazil	440444	440444	2.06%
4	United States	414028	414028	0.03%
5	India	189412	189412	2.04%
6	Argentina	183528	183528	4.00%
7	France	131516	131516	0.21%
8	Mexico	118506	118506	0.73%
9	Taiwan (, Province Of China)	113333	113333	0.61%
10	Japan	94600	94600	0.06%
11	Chile	84345	84345	2.13%
12	Spain	81771	81771	0.41%
13	Germany	65753	65753	0.10%
14	United Kingdom	56343	56343	0.07%
15	Canada	52297	52297	0.07%
16	Turkey	50474	50474	0.61%
17	Israel	48023	48023	1.19%
18	Malaysia	43100	43100	1.38%
19	Peru	41767	41767	4.08%
20	Italy	37554	37554	0.16%
21	Poland	29708	29708	0.25%
22	Colombia	29591	29591	1.06%
23	Portugal	27562	27562	0.81%
24	Thailand	26250	26250	0.72%
25	Hong Kong	25575	25575	0.37%
26	Austria	24025	24025	0.35%
27	Russian Federation	22974	22974	0.15%
28	Singapore	22960	22960	0.61%
29	Netherlands	17632	17632	0.09%
30	Australia	15444	15444	0.05%
31	Switzerland	13912	13912	0.19%
32	(unknown)	12721	12721	n/a
33	Hungary	12121	12121	0.38%
34	Venezuela	11990	11990	0.35%
35	Philippines	11591	11591	0.50%
36	Sweden	11557	11557	0.07%
37	Belgium	10849	10849	0.20%
38	Pakistan	10330	10330	1.71%
39	Denmark	9758	9758	0.12%
40	European Union (can apply to any country in Europe)	8447	8447	0.01%
41	Romania	7870	7870	0.15%
42	Morocco	7328	7328	1.25%
43	Czech Republic	7059	7059	0.15%
44	Greece	6073	6073	0.24%
45	South Africa	5664	5664	0.05%
46	Uruguay	5655	5655	1.52%
47	Finland	5626	5626	0.07%
48	Norway	5274	5274	0.08%
49	Indonesia	4592	4592	0.21%
50	Slovenia	4502	4502	0.43%

**CBL**

rank	country	entries	range	quota
1	Brazil	680678	680678	3,19%
2	China	650290	650290	0,54%
3	United States	388390	388390	0,03%
4	India	288320	288320	3,10%
5	Turkey	224254	224254	2,72%
6	Germany	223394	223394	0,32%
7	Poland	219170	219170	1,83%
8	Korea, Republic of (South)	211228	211228	0,38%
9	Russian Federation	148319	148319	0,95%
10	Viet Nam	140113	140113	4,02%
11	Mexico	119589	119589	0,74%
12	France	115280	115280	0,18%
13	Taiwan (, Province Of China	113647	113647	0,61%
14	Argentina	104760	104760	2,29%
15	Thailand	104726	104726	2,89%
16	Israel	87865	87865	2,18%
17	United Kingdom	84331	84331	0,10%
18	Chile	83818	83818	2,11%
19	Spain	73159	73159	0,37%
20	Egypt	71019	71019	3,38%
21	Italy	58064	58064	0,24%
22	Morocco	52253	52253	8,92%
23	Romania	48982	48982	0,91%
24	Peru	48435	48435	4,73%
25	Portugal	46345	46345	1,36%
26	Colombia	46321	46321	1,66%
27	Philippines	45585	45585	1,96%
28	Hungary	40805	40805	1,27%
29	Ukraine	39721	39721	1,64%
30	Malaysia	36058	36058	1,15%
31	Japan	35516	35516	0,02%
32	Iran, Islamic Republic of	30651	30651	2,29%
33	Indonesia	24545	24545	1,12%
34	Australia	22872	22872	0,07%
35	Hong Kong	22284	22284	0,32%
36	Croatia (Hrvatska)	20563	20563	2,30%
37	European Union (can apply to any country in Europe)	20064	20064	0,02%
38	Czechoslovakia (former)	19806	19806	2,98%
39	Pakistan	19164	19164	3,18%
40	Greece	17841	17841	0,70%
41	Saudi Arabia	15950	15950	1,06%
42	Sweden	14348	14348	0,09%
43	Canada	14193	14193	0,02%
44	Switzerland	13782	13782	0,19%
45	Uruguay	13251	13251	3,57%
46	Slovakia (Slovak Republic)	13216	13216	0,98%
47	Bulgaria	12275	12275	0,52%
48	Dominican Republic	11849	11849	5,08%
49	Slovenia	11148	11148	1,05%
50	Algeria	11133	11133	4,12%

**pbl.spamhaus.org**

rank	country	entries	range	quota
1	United States	49604	120467378	8.78%
2	Japan	5999	28940095	18.74%
3	China	8383	27448962	23.43%
4	Germany	1418	23568477	34.17%
5	(unknown)	3078	16897301	n/a
6	Canada	9233	10427689	14.29%
7	United Kingdom	2458	7778451	12.12%
8	France	1794	6940961	38.63%
9	Taiwan (, Province Of China)	1259	6923462	37.01%
10	Mexico	848	6313481	38.83%
11	Spain	925	6247749	31.16%
12	Korea, Republic of (South)	3595	5944359	10.73%
13	Italy	976	5037499	20.90%
14	Brazil	4497	4405759	20.68%
15	Poland	2077	2916732	24.29%
16	Turkey	373	2730352	14.25%
17	Netherlands	1702	2660048	33.09%
18	European Union (can apply to any country in Europe)	2408	2630553	2.18%
19	Sweden	649	2507387	15.32%
20	Australia	3330	2476309	7.61%
21	India	1067	2374248	25.57%
22	Israel	473	2049958	50.85%
23	Switzerland	253	1201165	47.08%
24	Venezuela	496	1181584	16.76%
25	Austria	3401	1053819	34.17%
26	Thailand	626	1028181	31.34%
27	Viet Nam	95	1018876	28.36%
28	Singapore	553	985188	26.27%
29	Russian Federation	2720	948582	17.61%
30	Chile	460	921708	6.05%
31	Argentina	2520	905135	23.76%
32	Hong Kong	189	854910	19.74%
33	Hungary	949	790364	12.38%
34	South Africa	689	787258	12.30%
35	Belgium	328	721306	24.60%
36	Finland	553	687479	7.28%
37	Denmark	1074	644967	8.02%
38	Colombia	346	639495	23.44%
39	Greece	679	616542	8.22%
40	Romania	456	570023	24.20%
41	Malaysia	590	537889	24.38%
42	Czech Republic	732	507947	17.18%
43	Norway	579	500592	10.66%
44	Portugal	186	444212	7.58%
45	Morocco	87	434254	9.24%
46	Philippines	389	394928	74.11%
47	Ukraine	587	392383	17.02%
48	New Zealand (Aotearoa)	378	335076	6.75%
49	Egypt	347	281358	13.40%
50	Slovakia (Slovak Republic)	69	277188	20.44%

**xbl.spamhaus.org**

rank	country	entries	range	quota
1	China	986446	986446	0.83%
2	Brazil	734597	734597	3.64%
3	United States	393603	393603	0.03%
4	India	302376	302376	3.72%
5	Korea, Republic of (South)	273741	273741	0.50%
6	Turkey	229770	229770	2.80%
7	Germany	229639	229639	0.33%
8	Poland	220622	220622	1.85%
9	Russian Federation	150312	150312	0.96%
10	Viet Nam	138570	138570	3.99%
11	Taiwan (, Province Of China	129607	129607	2.91%
12	Argentina	124993	124993	0.70%
13	Mexico	122637	122637	0.76%
14	France	118848	118848	0.19%
15	Thailand	106455	106455	2.96%
16	Chile	89821	89821	2.32%
17	Israel	89039	89039	2.21%
18	United Kingdom	85763	85763	0.10%
19	Spain	76340	76340	0.38%
20	Egypt	71184	71184	3.39%
21	Italy	60525	60525	0.25%
22	Morocco	53303	53303	9.24%
23	Colombia	50356	50356	1.84%
24	Romania	50071	50071	0.93%
25	Peru	49583	49583	4.88%
26	Portugal	47168	47168	1.39%
27	Philippines	46071	46071	2.00%
28	Hungary	41464	41464	1.30%
29	Ukraine	40075	40075	1.67%
30	Japan	38798	38798	1.25%
31	Malaysia	38554	38554	0.03%
32	Iran, Islamic Republic of	30901	30901	2.32%
33	Indonesia	24867	24867	1.14%
34	Hong Kong	24067	24067	0.36%
35	Australia	23339	23339	0.07%
36	European Union (can apply to any country in Europe)	21131	21131	0.02%
37	Croatia (Hrvatska)	20557	20557	2.31%
38	Czechoslovakia (former)	19901	19901	3.39%
39	Pakistan	19859	19859	3.00%
40	Greece	18413	18413	0.72%
41	Saudi Arabia	16117	16117	1.08%
42	Canada	15044	15044	0.02%
43	Sweden	14669	14669	0.09%
44	Switzerland	14058	14058	0.20%
45	Slovakia (Slovak Republic)	13810	13810	1.03%
46	Uruguay	13581	13581	3.70%
47	Bulgaria	12886	12886	0.55%
48	Dominican Republic	12672	12672	5.52%
49	Slovenia	11401	11401	4.31%
50	Algeria	11192	11192	1.08%

**dnswl.org**

rank	country	entries	range	quota
1	United States	5411	1392899	0,10%
2	European Union (can apply to any country in Europe)	548	799212	0,66%
3	Canada	320	329999	0,45%
4	Switzerland	559	190327	2,66%
5	Chile	7	65542	1,65%
6	United Kingdom	316	23126	0,03%
7	Germany	406	11540	0,02%
8	Hong Kong	36	8890	0,13%
9	Sweden	166	5331	0,03%
10	China	50	5211	0,00%
11	France	103	3990	0,01%
12	(unknown)	38	3353	n/a
13	Spain	252	3079	0,02%
14	Turkey	11	2121	0,03%
15	Venezuela	4	2051	0,06%
16	Italy	50	1927	0,01%
17	Netherlands	58	1906	0,01%
18	Australia	201	1769	0,01%
19	India	30	1560	0,02%
20	Belgium	27	1557	0,03%
21	Japan	24	1554	0,00%
22	Brazil	29	1063	0,01%
23	Israel	17	1047	0,03%
24	Russian Federation	27	823	0,01%
25	Austria	58	823	0,01%
26	Korea, Republic of (South)	16	796	0,00%
27	Singapore	26	613	0,02%
28	Denmark	46	587	0,01%
29	Taiwan (, Province Of China)	20	561	0,00%
30	Poland	26	536	0,00%
31	New Zealand (Aotearoa)	20	530	0,01%
32	South Africa	16	526	0,01%
33	Portugal	12	522	0,02%
34	Mexico	23	293	0,00%
35	Ireland	24	279	0,01%
36	Malaysia	16	271	0,01%
37	Latvia	6	261	0,02%
38	Estonia	4	259	0,03%
39	Morocco	2	257	0,04%
40	Togo	1	256	2,08%
41	Argentina	16	31	0,00%
42	Finland	12	27	0,00%
43	Czech Republic	10	25	0,00%
44	Norway	22	22	0,00%
45	Costa Rica	7	22	0,00%
46	Colombia	16	16	0,00%
47	United Arab Emirates	15	15	0,00%
48	Thailand	15	15	0,00%
49	Kenya	13	13	0,01%
50	Saudi Arabia	12	12	0,00%



**Union of all blacklists**

rank	country	entries	range	quota
1	United States	283463	121389419	8.72%
2	Japan	54708	29054339	18.81%
3	China	238876	27655748	22.88%
4	Germany	23064	23590123	34.20%
5	(unknown)	8599	17304975	n/a
6	Canada	40958	10656784	14.60%
7	United Kingdom	42859	7819107	9.47%
8	France	43123	6982290	10.88%
9	Taiwan (Province Of China)	16044	6931085	37.00%
10	Mexico	35877	6352605	39.07%
11	Spain	41962	6288530	31.36%
12	Korea, Republic of (South)	341145	6281771	11.33%
13	Italy	14789	5051822	20.96%
14	Brazil	90517	4491779	21.02%
15	Poland	18292	2932947	24.42%
16	European Union (can apply to any country)	18006	2777221	2.31%
17	Turkey	10519	2740753	33.21%
18	Netherlands	13867	2671966	13.88%
19	Sweden	13075	2520068	15.40%
20	Australia	22857	2497880	7.68%
21	India	36302	2376717	25.58%
22	Israel	17576	2067061	51.17%
23	Switzerland	6866	1207778	16.85%
24	Venezuela	5315	1185125	34.28%
25	Austria	9327	1059745	16.07%
26	Singapore	3328	1053498	28.08%
27	Thailand	9357	1036912	28.60%
28	Viet Nam	4937	1023718	29.38%
29	Russian Federation	37851	983203	6.27%
30	Argentina	48989	951094	20.74%
31	Chile	23935	945183	23.82%
32	South Africa	2343	920045	8.51%
33	Hong Kong	14319	869040	12.58%
34	Hungary	5895	795310	24.74%
35	Belgium	14992	735970	13.58%
36	Finland	2871	689797	8.04%
37	Colombia	13641	652790	23.34%
38	Denmark	7247	651140	8.30%
39	Greece	5532	621395	24.39%
40	Romania	28201	599043	11.09%
41	Malaysia	7705	545004	17.41%
42	Czech Republic	6777	513992	10.79%
43	Norway	4629	504642	7.65%
44	Portugal	8553	452579	13.26%
45	Morocco	1108	435275	74.28%
46	Philippines	12268	406297	17.51%
47	Ukraine	6085	397881	16.46%
48	New Zealand (Aotearoa)	3321	338019	6.81%
49	Egypt	7612	288623	13.74%
50	Slovakia (Slovak Republic)	1950	279069	20.58%

## E Annex - Blacklist entries by AS

The following pages show the blacklist entries assigned to Autonomous Systems. Each blacklist is on a separate page, listing the top 50 Autonomous Systems ordered by the sum of listed net ranges.

IP based DNSBLs that have been reviewed are:

- all.dnsbl.sorbs.net
- UCEPROTECT - Level 1
- NiX Spam
- dnsbl.ahbl.org
- sbl.spamhaus.org
- dnsbl.njabl.org
- CBL
- pbl.spamhaus.org
- xbl.spamhaus.org
- dnswl.org
- union of all blacklists

For the origin of the data please see chapter 7.2.

The tables list amounts by AS. *Rank* indicated the position of the country ordered by the sum of listed net ranges. *asid* describes the identifying number of the Autonomous System, the name is given in column *name*. *Entries* indicates the number of net range entries in the table. Contrarily *Range* sums the size of all net ranges and gives information about how many particular hosts (single IP addresses) have been listed. *Quota* gives information about the percentage of the covered range of an AS, i.e. which ratio of the amount of a Autonomous System's assigned IP addresses is listed in a blacklist.

**all.dnsbl.sorbs.net**

rank	asid	name	entries	range	quota
1	4837	CHINA169-BACKBONE CNCGROUP China169 Backbon	138044	138044	0,68%
2	9121	TTNET TTnet Autonomous System	127066	127066	1,30%
3	4766	KIXS-AS-KR Korea Telecom	124341	124341	0,58%
4	577	BACOM - Bell Canada	9984	75519	2,11%
5	11784	ASN-AV8 - Plain Aivation, Inc	2	67584	100,00%
6	3269	ASN-IBSNAZ TELECOM ITALIA	67311	67311	0,57%
7	11188	LACOUNTY-ISD - LOS ANGELES COUNTY - INTERNA	72	65607	33,24%
8	35921	IFCI-US - InternetFCI LLC	1	65536	83,39%
9	3496	Maraven	1	65536	31,92%
10	9829	BSNL-NIB National Internet Backbone	61632	61632	3,15%
11	3352	TELEFONICA-DATA-ESPANA Internet Access Network	57138	57138	0,71%
12	3320	DTAG Deutsche Telekom AG	55654	55654	0,22%
13	7018	ATT-INTERNET4 - AT&T WorldNet Services	4640	53786	0,01%
14	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELES	46736	46736	1,94%
15	22927	Telefonica de Argentina	38636	38636	4,98%
16	3215	AS3215 France Telecom - Orange	33298	33298	0,34%
17	5617	TPNET Polish Telecom's commercial IP network	32409	32409	0,82%
18	2856	BT-UK-AS BTnet UK Regional network	23172	23172	0,21%
19	6713	IAM-AS	22933	22933	1,75%
20	4812	CHINANET-SH-AP China Telecom (Group)	21445	21445	0,30%
21	4814	CHINA169-BBN CNCGROUP IP network	21102	21102	0,88%
22	7643	VNN-AS-AP Vietnam Posts and Telecommunications (V	20852	20852	2,15%
23	3209	Arcor IP-Network	20473	20473	0,60%
24	8167	TELESC - Telecomunicacoes de Santa Catarina SA	19693	19693	0,92%
25	3462	HINET Data Communication Business Group	19608	19608	0,25%
26	17858	KRNIC-ASBLOCK-AP KRNIC	19147	19147	0,43%
27	4788	TMNET-AS-AP TM Net, Internet Service Provider	17491	17491	0,37%
28	5430	FREENETDE freenet Cityline GmbH	15261	15261	0,46%
29	3356	LEVEL3 Level 3 Communications	14683	14683	0,01%
30	7132	SBIS-AS - AT&T Internet Services	14644	14644	0,05%
31	18101	RIL-IDC Reliance Infocom Ltd Internet Data Centre,	13789	13789	1,68%
32	13184	HANSENET HanseNet Telekommunikation GmbH	13758	13758	1,16%
33	12479	UNI2-AS Uni2 Autonomous System	12892	12892	0,37%
34	6739	ONO-AS Cableuropa - ONO	12745	12745	0,84%
35	9299	IPG-AS-AP Philippine Long Distance Telephone Compar	12662	12662	1,41%
36	1668	AOL-ATDN - AOL Transit Data Network	11792	11792	0,09%
37	16338	AUNA_TELECOM-AS AUNA Autonomous System	11212	11212	0,80%
38	209	ASN-QWEST - Qwest	9867	9867	0,01%
39	8551	BEZEQ-INTERNATIONAL-AS Bezeqint Internet Backbor	9845	9845	0,77%
40	8452	TEDATA TEDATA	9755	9755	0,91%
41	4230	Embratel	9279	9279	0,25%
42	7552	VIETEL-AS-AP Viettel Corporation	8881	8881	0,20%
43	5462	CABLEINET Telewest Broadband	8805	8805	0,32%
44	17839	DREAMPLUS-AS-KR DreamcityMedia	8600	8600	3,09%
45	15557	LDCOMNET NEUF CEGETEL (formerly LDCOM NETW	8585	8585	0,28%
46	9394	CRNET CHINA RAILWAY Internet(CRNET)	8441	8441	0,07%
47	18403	FPT-AS-AP The Corporation for Financing & Promoting T	8125	8125	2,65%
48	12322	PROXAD AS for Proxad/Free ISP	8063	8063	0,09%
49	9105	TISCALI-UK Tiscali UK	7895	7895	0,33%
50	5483	HTC-AS Hungarian Telecom	7844	7844	1,24%

**UCEPROTECT - Level 1**

rank	asid	name	entries	range	quota
1	5617	TPNET Polish Telecom's commercial IP network	46715	46715	1.18%
2	9121	TTNET TTnet Autonomous System	31120	31120	0.32%
3	3320	DTAG Deutsche Telekom AG	22613	22613	0.09%
4	3462	HINET Data Communication Business Group	17041	17041	0.22%
5	4837	CHINA169-BACKBONE CNCGROUP China169 Backbo	16386	16386	0.08%
6	3215	AS3215 France Telecom - Orange	15292	15292	0.16%
7	6713	IAM-AS	12680	12680	0.97%
8	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELESP	12583	12583	0.52%
9	4766	KIXS-AS-KR Korea Telecom	11426	11426	0.05%
10	9829	BSNL-NIB National Internet Backbone	11025	11025	0.56%
11	8167	TELESC - Telecomunicacoes de Santa Catarina SA	9712	9712	0.45%
12	7643	VNN-AS-AP Vietnam Posts and Telecommunications (VI	8187	8187	0.85%
13	22927	Telefonica de Argentina	7448	7448	0.96%
14	3269	ASN-IBSNAZ TELECOM ITALIA	7297	7297	0.06%
15	3209	Arcor IP-Network	7261	7261	0.21%
16	4788	TMNET-AS-AP TM Net. Internet Service Provider	7058	7058	0.15%
17	8452	TEDATA TEDATA	7019	7019	0.65%
18	8359	COMSTAR-Direct Moscow region network	6477	6477	0.91%
19	15557	LDCOMNET NEUF CEGETEL	6373	6373	0.21%
20	9737	TOTNET-TH-AS-AP Telephone Org. of Thailand	5808	5808	0.67%
21	8551	BEZEQ-INTERNATIONAL-AS Bezeqint Internet Backb.	5714	5714	0.45%
22	3352	TELEFONICA-DATA-ESPANA Internet Access TDE	5340	5340	0.07%
23	9299	IPG-AS-AP Philippine Long Distance Telephone Comp.	4825	4825	0.54%
24	209	ASN-QWEST - Qwest	4790	4790	0.00%
25	2856	BT-UK-AS BTnet UK Regional network	4203	4203	0.04%
26	5430	FREENETDE freenet Cityline GmbH	4082	4082	0.12%
27	24863	LINKdotNET-AS	3648	3648	1.05%
28	4812	CHINANET-SH-AP China Telecom (Group)	3620	3620	0.05%
29	6830	UPC UPC Broadband	3601	3601	0.06%
30	18101	RIL-IDC Reliance Infocom Ltd Internet Data Centre.	3465	3465	0.42%
31	6849	UKRTELNET JSC UKRTELECOM.	3345	3345	2.70%
32	7132	SBIS-AS - AT&T Internet Services	3215	3215	0.01%
33	12876	AS12876 Telecom Italia France	2818	2818	0.11%
34	11427	SCRR-11427 - Road Runner HoldCo LLC	2739	2739	0.09%
35	13184	HANSENET HanseNet Telekommunikation GmbH	2688	2688	0.23%
36	9050	RTD RTD-ROMTELECOM ASN	2677	2677	0.59%
37	12741	INTERNETIA-AS Netia SA	2647	2647	0.40%
38	11351	RR-NYSREGION-ASN-01 - Road Runner HoldCo LLC	2534	2534	0.06%
39	20858	EGYNET-AS	2494	2494	1.51%
40	5089	NTL NTL Group Limited	2477	2477	0.02%
41	9116	GOLDENLINES-ASN Golden Lines Main AS	2386	2386	0.19%
42	6167	CELLCO-PART - Cellco Partnership	2376	2376	0.01%
43	17839	DREAMPLUS-AS-KR DreamcityMedia	2374	2374	0.85%
44	5486	SMILE-ASN Euronet Digital Communications. Israel	2359	2359	0.18%
45	8228	CEGETEL-AS CEGETEL ENTREPRISES	2311	2311	0.12%
46	5713	SAIX-NET	2298	2298	0.15%
47	5483	HTC-AS Hungarian Telecom	2222	2222	0.35%
48	5462	CABLEINET Telewest Broadband	2222	2222	0.08%
49	11426	SCRR-11426 - Road Runner HoldCo LLC	2190	2190	0.08%
50	9105	TISCALI-UK Tiscali UK	2167	2167	0.09%

**NiX Spam**

rank	asid	name	entries	range	quota
1	9121	TTNET TNet Autonomous System	3104	3104	0,03%
2	3320	DTAG Deutsche Telekom AG	3056	3056	0,01%
3	4766	KIXS-AS-KR Korea Telecom	1821	1821	0,01%
4	3215	AS3215 France Telecom - Orange	1440	1440	0,02%
5	4837	CHINA169-BACKBONE CNCGROUP China169 Backbone	1429	1429	0,01%
6	5617	TPNET Polish Telecom's commercial IP network	1404	1404	0,04%
7	3209	Arcor IP-Network	800	800	0,02%
8	3269	ASN-IBSNAZ TELECOM ITALIA	778	778	0,01%
9	8359	COMSTAR COMSTAR-Direct Moscow region network	724	724	0,10%
10	8167	TELESC - Telecomunicacoes de Santa Catarina SA	677	677	0,03%
11	6713	IAM-AS	666	666	0,05%
12	8551	BEZEQ-INTERNATIONAL-AS Bezeqint Internet Backbone	645	645	0,05%
13	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELESP	636	636	0,03%
14	6830	UPC UPC Broadband	634	634	0,01%
15	15557	LDCOMNET NEUF CEGETEL (formerly LDCOM NETWORK)	557	557	0,02%
16	3352	TELEFONICA-DATA-ESPANA Internet Access Network	555	555	0,01%
17	2856	BT-UK-AS BTnet UK Regional network	545	545	0,01%
18	22927	Telefonica de Argentina	533	533	0,07%
19	17858	KRNIC-ASBLOCK-AP KRNIC	524	524	0,01%
20	3462	HINET Data Communication Business Group	501	501	0,01%
21	5089	NTL NTL Group Limited	497	497	0,00%
22	20115	CHARTER-NET-HKY-NC - Charter Communications	469	469	0,02%
23	5462	CABLEINET Telewest Broadband	434	434	0,02%
24	5430	FREENETDE freenet Cityline GmbH	430	430	0,01%
25	9829	BSNL-NIB National Internet Backbone	413	413	0,02%
26	209	ASN-QWEST - Qwest	408	408	0,00%
27	7132	SBIS-AS - AT&T Internet Services	382	382	0,00%
28	11351	RR-NYSREGION-ASN-01 - Road Runner HoldCo LLC	381	381	0,01%
29	4814	CHINA169-BBN CNCGROUP IP network	359	359	0,02%
30	7643	VNN-AS-AP Vietnam Posts and Telecommunications (VPT)	354	354	0,04%
31	4788	TMNET-AS-AP TM Net, Internet Service Provider	341	341	0,01%
32	13184	HANSENET HanseNet Telekommunikation GmbH	322	322	0,03%
33	9299	IPG-AS-AP Philippine Long Distance Telephone Company	312	312	0,04%
34	4812	CHINANET-SH-AP China Telecom (Group)	307	307	0,00%
35	6739	ONO-AS Cableuropa - ONO	296	296	0,02%
36	9737	TOTNET-TH-AS-AP Telephone Organization of Thailand	287	287	0,03%
37	8402	CORBINA-AS Corbina telecom	284	284	0,07%
38	4230	Embratel	280	280	0,01%
39	16338	AUNA_TELECOM-AS AUNA Autonomous System	279	279	0,02%
40	11427	SCRR-11427 - Road Runner HoldCo LLC	274	274	0,01%
41	9116	GOLDENLINES-ASN Golden Lines Main Autonomous System	273	273	0,02%
42	11426	SCRR-11426 - Road Runner HoldCo LLC	240	240	0,01%
43	8881	VERSATEL Versatel Global Network	239	239	0,02%
44	8452	TEDATA TEDATA	234	234	0,02%
45	6805	TDDE-ASN1 Telefonica Deutschland Autonomous System	219	219	0,01%
46	8997	ASN-SPBNIT SPBNIT-RU Autonomous System	210	210	0,09%
47	6400	VERIZON DOMINICANA	208	208	0,11%
48	7015	CCCH-AS2 - Comcast Cable Communications Holdings, Inc.	204	204	0,00%
49	21502	ASN-NUMERICABLE NUMERICABLE is a cabled network	202	202	0,01%
50	22291	CHARTER-LA - Charter Communications	200	200	0,02%

**dnsbl.ahbl.org**

rank	asid	name	entries	range	quota
1	4837	CHINA169-BACKBONE CNCGROUP China169 Backbon	202480	202480	0,99%
2	4766	KIXS-AS-KR Korea Telecom	165025	165025	0,77%
3	7132	SBIS-AS - AT&T Internet Services	90526	91037	0,30%
4	3462	HINET Data Communication Business Group	61304	61304	0,79%
5	22927	Telefonica de Argentina	52871	52871	6,81%
6	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELESP	51566	51566	2,14%
7	4812	CHINANET-SH-AP China Telecom (Group)	51564	51564	0,73%
8	3215	AS3215 France Telecom - Orange	47886	47886	0,49%
9	4814	CHINA169-BBN CNCGROUP IP network	43677	43677	1,81%
10	9121	TTNET TTnet Autonomous System	41845	42100	0,43%
11	3320	DTAG Deutsche Telekom AG	39259	39259	0,16%
12	9829	BSNL-NIB National Internet Backbone	37466	37466	1,92%
13	4788	TMNET-AS-AP TM Net, Internet Service Provider	37127	37127	0,78%
14	16338	AUNA_TELECOM-AS AUNA Autonomous System	13803	30186	2,15%
15	1668	AOL-ATDN - AOL Transit Data Network	29231	29231	0,21%
16	3269	ASN-IBSNAZ TELECOM ITALIA	27511	27766	0,24%
17	8167	TELESC - Telecomunicacoes de Santa Catarina SA	26154	26154	1,22%
18	9394	CRNET CHINA RAILWAY Internet(CRNET)	21284	21284	0,17%
19	3352	TELEFONICA-DATA-ESPANA Internet Access Network	20825	21080	0,26%
20	3356	LEVEL3 Level 3 Communications	18932	18932	0,02%
21	5617	TPNET Polish Telecom's commercial IP network	18831	18831	0,48%
22	12322	PROXAD AS for Proxad/Free ISP	17895	17895	0,20%
23	4230	Embratel	16515	16515	0,45%
24	2856	BT-UK-AS BTnet UK Regional network	13846	13846	0,13%
25	3786	LGDACOM LG DACOM Corporation	13670	13670	0,11%
26	20115	CHARTER-NET-HKY-NC - Charter Communications	12753	12753	0,42%
27	9506	MAGIX-SG-AP Magix Broadband Network	12533	12533	2,69%
28	6830	UPC UPC Broadband	11515	11515	0,20%
29	18101	RIL-IDC Reliance Infocom Ltd Internet Data Centre,	11232	11232	1,37%
30	15557	LDCOMNET NEUF CEGETEL (formerly LDCOM NETW	10566	10566	0,35%
31	9304	HUTCHISON-AS-AP Hutchison Global Communications	10273	10273	0,43%
32	577	BACOM - Bell Canada	10158	10158	0,28%
33	3209	Arcor IP-Network	9997	9997	0,30%
34	5462	CABLEINET Telewest Broadband	9900	9900	0,36%
35	6739	ONO-AS Cableuropa - ONO	8283	8283	0,54%
36	29761	OC3-NETWORKS-AS-NUMBER - OC3 Networks & Web	8	8199	50,04%
37	7018	ATT-INTERNET4 - AT&T WorldNet Services	7803	8058	0,00%
38	12542	TVCABO Autonomous System	8023	8023	1,15%
39	17676	JPNIC-JP-ASN-BLOCK Japan Network Information Cent	7706	7706	0,01%
40	16735	Companhia de Telecomunicacoes do Brasil Central	7356	7356	2,99%
41	1221	ASN-TELSTRA Telstra Pty Ltd	7192	7192	0,01%
42	4670	HYUNDAI-KR Shinbiro	7096	7096	0,47%
43	8551	BEZEQ-INTERNATIONAL-AS Bezeqint Internet Backbor	6870	6870	0,54%
44	6713	IAM-AS	6864	6864	0,53%
45	13184	HANSENET HanseNet Telekommunikation GmbH	6517	6517	0,55%
46	10091	SCV-AS-AP SCV Broadband Access Provider	6411	6411	0,56%
47	209	ASN-QWEST - Qwest	6024	6279	0,00%
48	12876	AS12876 Telecom Italia France	6229	6229	0,25%
49	19548	ADELPHIA-AS2 - Road Runner HoldCo LLC	6169	6169	0,13%
50	2056	AOL-AS - America Online	6112	6112	0,71%

**sbl.spamhaus.org**

rank	asid	name	entries	range	quota
1	18101	RIL-IDC Reliance Infocom Ltd Internet Data Centre,	7	32773	4.08%
2	4837	CHINA169-BACKBONE CNCGROUP China169 Backbon	53	18345	0.09%
3	36114	RDTECH-ASN - R & D Technologies, LLC	4	12321	50.13%
4	35709	PLAZA Internet Service Provider	1	8192	33.33%
5	2828	XO-AS15 - XO Communications	29	8130	0.11%
6	29614	GHANATEL-AS	5	5120	11.17%
7	3257	TISCALI-BACKBONE Tiscali Intl Network BV	13	4874	0.00%
8	17858	KRNIC-ASBLOCK-AP KRNIC	8	4613	0.10%
9	174	COGENT Cogent/PSI	6	4172	0.01%
10	35153	RELIANS Autonomous System	1	4096	50.00%
11	19403	TECHALLIANCEGROUP - TECHNOLOGY ALLIANCE G	1	4096	64.00%
12	3790	RADIGRAFICA COSTARRICENSE	1	4096	2.88%
13	9116	GOLDENLINES-ASN Golden Lines Main Autonomous S	28	3760	0.30%
14	8346	SONATEL-AS Autonomous System	16	3143	4.48%
15	15756	CARAVAN ISP "CARAVAN"	8	3075	6.26%
16	7552	VIETEL-AS-AP Viettel Corporation	7	3075	0.07%
17	4808	CHINA169-BJ CNCGROUP IP network China169 Beijing	12	2825	0.06%
18	17968	DQTNET Daqing zhongji petroleum telecommunication c	6	2818	1.49%
19	19318	NJIX-AS-1 - NEW JERSEY INTERNATIONAL INTERNE	3	2576	9.12%
20	22927	Telefonica de Argentina	13	2183	0.28%
21	15227	WVFIBERNET - FiberNet of West Virginia	5	2081	1.51%
22	16338	AUNA_TELECOM-AS AUNA Autonomous System	16	2060	0.15%
23	11486	WAN - Worldcom Advance Networks	12	2058	0.37%
24	7132	SBIS-AS - AT&T Internet Services	34	2055	100.00%
25	1257	TELE2	8	2055	0.04%
26	2116	ASN-CATCHCOM Catch Communications	3	2050	0.28%
27	42119	BALSAX-AS BalSax-IT ApS AS Number	1	2048	100.00%
28	35935	MARKETEXP-1 - Ecommerce Marketing Consultants, LL	1	2048	100.00%
29	42461	PLANETA-AS ISP Planeta	1	2048	25.00%
30	13488	CBWU-13488 - Continental Broadband Florida, Inc DBA	15	1859	1.26%
31	17897	CHINATELECOM-HLJ-AS-AP asn for Heilongjiang Provi	4	1792	0.36%
32	6939	HURRICANE - Hurricane Electric	24	1773	0.08%
33	25847	SERVINT - ServInt Corporation	11	1697	2.38%
34	12491	IPPLANET-AS IPPlanet	20	1691	1.30%
35	10013	JPNIC-NET-JP-AS-BLOCK Japan Network Information C	23	1600	0.27%
36	11194	NUNETPA - NuNet Inc	3	1536	3.75%
37	10439	CARI - San Diego Commercial Internet Exchange	8	1411	1.32%
38	1239	SPRINTLINK - Sprint	11	1351	0.00%
39	36158	DEV8E - Dev8 Entertainment	2	1280	100.00%
40	17676	JPNIC-JP-ASN-BLOCK Japan Network Information Cent	13	1225	0.00%
41	3462	HINET Data Communication Business Group	17	1186	0.02%
42	6830	UPC UPC Broadband	26	1116	0.02%
43	9488	SNU-AS-KR Seoul National University	4	1027	100.00%
44	12576	ORANGE-PCS Orange PCS Limited	1	1024	0.20%
45	41731	NEVSKCC-AS NEVAICON LTD	1	1024	100.00%
46	33520	GAMUT-HOSTING - GAMUT HOSTING	1	1024	14.29%
47	13999	MegaCable SA de CV	1	1024	0.09%
48	9391	UNSPECIFIED GUANGDONG HIGHWAY BROADBAND	2	1024	3.13%
49	31159	NETCATHOST-AS NetcatHosting	1	1024	100.00%
50	33775	NITEL-AS	4	1024	100.00%

**dnsbl.njabl.org**

rank	asid	name	entries	range	quota
1	4837	CHINA169-BACKBONE CNCGROUP China169 Backbon	283199	283199	1,39%
2	4766	KIXS-AS-KR Korea Telecom	192823	192823	0,89%
3	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELESP	165597	165597	6,87%
4	7132	SBIS-AS - AT&T Internet Services	126420	126420	0,41%
5	22927	Telefonica de Argentina	100363	100363	12,93%
6	3462	HINET Data Communication Business Group	84377	84377	1,09%
7	3215	AS3215 France Telecom - Orange	73138	73138	0,74%
8	9829	BSNL-NIB National Internet Backbone	71128	71128	3,64%
9	4812	CHINANET-SH-AP China Telecom (Group)	68930	68930	0,97%
10	8167	TELESC - Telecomunicacoes de Santa Catarina SA	68829	68829	3,21%
11	4814	CHINA169-BBN CNCGROUP IP network	63749	63749	2,64%
12	9121	TTNET TTnet Autonomous System	48181	48181	0,49%
13	4788	TMNET-AS-AP TM Net, Internet Service Provider	41861	41861	0,88%
14	9394	CRNET CHINA RAILWAY Internet(CRNET)	31345	31345	0,25%
15	4230	Embratel	30188	30188	0,82%
16	3269	ASN-IBSNAZ TELECOM ITALIA	29143	29143	0,25%
17	3352	TELEFONICA-DATA-ESPANA Internet Access Network	27296	27296	0,34%
18	16338	AUNA_TELECOM-AS AUNA Autonomous System	24049	24049	1,72%
19	3320	DTAG Deutsche Telekom AG	21377	21377	0,09%
20	5617	TPNET Polish Telecom's commercial IP network	20733	20733	0,53%
21	3786	LGDACOM LG DACOM Corporation	20694	20694	0,17%
22	3356	LEVEL3 Level 3 Communications	19912	19912	0,02%
23	12322	PROXAD AS for Proxad/Free ISP	19909	19909	0,22%
24	2856	BT-UK-AS BTnet UK Regional network	19556	19556	0,18%
25	18101	RIL-IDC Reliance Infocom Ltd Internet Data Centre,	18570	18570	2,27%
26	6830	UPC UPC Broadband	17454	17454	0,30%
27	20115	CHARTER-NET-HKY-NC - Charter Communications	17101	17101	0,56%
28	16735	Companhia de Telecomunicacoes do Brasil Central	15158	15158	6,17%
29	9304	HUTCHISON-AS-AP Hutchison Global Communications	14266	14266	0,59%
30	9506	MAGIX-SG-AP Magix Broadband Network	14113	14113	3,03%
31	4670	HYUNDAI-KR Shinbiro	13668	13668	0,90%
32	6739	ONO-AS Cableuropa - ONO	13068	13068	0,86%
33	5462	CABLEINET Telewest Broadband	12623	12623	0,45%
34	17676	JPNIC-JP-ASN-BLOCK Japan Network Information Cent	12521	12521	0,02%
35	3209	Arcor IP-Network	11373	11373	0,34%
36	8551	BEZEQ-INTERNATIONAL-AS Bezeqint Internet Backbor	11347	11347	0,89%
37	5089	NTL NTL Group Limited	11226	11226	0,09%
38	12542	TVCABO Autonomous System	10516	10516	1,51%
39	19548	ADELPHIA-AS2 - Road Runner HoldCo LLC	10385	10385	0,22%
40	8404	CABLECOM Cablecom GmbH	9162	9162	0,97%
41	7018	ATT-INTERNET4 - AT&T WorldNet Services	8876	8876	0,00%
42	9116	GOLDENLINES-ASN Golden Lines Main Autonomous Sy	8427	8427	0,68%
43	13184	HANSENET HanseNet Telekommunikation GmbH	8222	8222	0,69%
44	15557	LDCOMNET NEUF CEGETEL (formerly LDCOM NETW	7807	7807	0,26%
45	3292	TDC TDC Data Networks	7656	7656	0,14%
46	5483	HTC-AS Hungarian Telecom	7630	7630	1,21%
47	1221	ASN-TELSTRA Telstra Pty Ltd	7434	7434	0,01%
48	18881	Global Village Telecom	7427	7427	1,24%
49	5486	SMILE-ASN Euronet Digital Communications, (1992) LTD	7352	7352	0,55%
50	6713	IAM-AS	7328	7328	0,56%



**CBL**

rank	asid	name	entries	range	quota
1	9121	TTNET TNet Autonomous System	222370	222370	2,27%
2	4837	CHINA169-BACKBONE CNCGROUP China169 Backbon	199929	199929	0,98%
3	8167	TELESC - Telecomunicacoes de Santa Catarina SA	189030	189030	8,82%
4	5617	TPNET Polish Telecom's commercial IP network	187814	187814	4,75%
5	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELESP	173487	173487	7,20%
6	9829	BSNL-NIB National Internet Backbone	132433	132433	6,77%
7	3320	DTAG Deutsche Telekom AG	123482	123482	0,50%
8	7643	VNN-AS-AP Vietnam Posts and Telecommunications (V	103026	103026	10,64%
9	3462	HINET Data Communication Business Group	92033	92033	1,19%
10	22927	Telefonica de Argentina	73944	73944	9,53%
11	4766	KIXS-AS-KR Korea Telecom	63392	63392	0,29%
12	6713	IAM-AS	52204	52204	3,99%
13	3215	AS3215 France Telecom - Orange	47208	47208	0,48%
14	3269	ASN-IBSNAZ TELECOM ITALIA	45189	45189	0,38%
15	3209	Arcor IP-Network	42887	42887	1,27%
16	9737	TOTNET-TH-AS-AP Telephone Organization of Thailand	42285	42285	4,86%
17	4788	TMNET-AS-AP TM Net, Internet Service Provider	34304	34304	0,72%
18	3352	TELEFONICA-DATA-ESPANA Internet Access Network	32865	32865	0,41%
19	8551	BEZEQ-INTERNATIONAL-AS Bezeqint Internet Backbor	29679	29679	2,32%
20	4812	CHINANET-SH-AP China Telecom (Group)	29549	29549	0,42%
21	18101	RIL-IDC Reliance Infocom Ltd Internet Data Centre,	28435	28435	3,47%
22	4814	CHINA169-BBN CNCGROUP IP network	26967	26967	1,12%
23	5483	HTC-AS Hungarian Telecom	23119	23119	3,66%
24	6849	UKRTELNET JSC UKRTELECOM,	23115	23115	18,66%
25	18881	Global Village Telecom	23028	23028	3,85%
26	209	ASN-QWEST - Qwest	21361	21361	0,02%
27	8452	TEDATA TEDATA	20331	20331	1,89%
28	8359	COMSTAR COMSTAR-Direct Moscow region network	20199	20199	2,83%
29	9105	TISCALI-UK Tiscali UK	20131	20131	0,83%
30	2856	BT-UK-AS BTnet UK Regional network	19302	19302	0,17%
31	15557	LDCOMNET NEUF CEGETEL (formerly LDCOM NETW	19100	19100	0,62%
32	9299	IPG-AS-AP Philippine Long Distance Telephone Compar	18971	18971	2,12%
33	13184	HANSENET HanseNet Telekommunikation GmbH	18619	18619	1,57%
34	17974	TELKOMNET-AS2-AP PT TELEKOMUNIKASI INDONESIA	17957	17957	2,49%
35	7132	SBIS-AS - AT&T Internet Services	17202	17202	0,06%
36	18403	FPT-AS-AP The Corporation for Financing & Promoting T	16871	16871	5,49%
37	12876	AS12876 Telecom Italia France	16671	16671	0,68%
38	16735	Companhia de Telecomunicacoes do Brasil Central	16142	16142	6,57%
39	15475	NOL	16128	16128	2,39%
40	5391	T-HT T-Com Croatia Internet network	16009	16009	3,24%
41	4230	Embratel	15313	15313	0,42%
42	4808	CHINA169-BJ CNCGROUP IP network China169 Beijing	15042	15042	0,29%
43	9304	HUTCHISON-AS-AP Hutchison Global Communications	14700	14700	0,61%
44	5486	SMILE-ASN Euronet Digital Communications, (1992) LTD	14488	14488	1,09%
45	19429	ETB - Colombia	14421	14421	4,34%
46	9050	RTD RTD-ROMTELECOM Autonomous System Number	14266	14266	3,15%
47	20858	EGYNET-AS	13973	13973	8,46%
48	7552	VIETEL-AS-AP Vietel Corporation	13848	13848	0,32%
49	24863	LINKdotNET-AS	13618	13618	3,92%
50	6057	Administracion Nacional de Telecomunicaciones	12985	12985	2,60%

**pbl.spamhaus.org**

rank	asid	name	entries	range	quota
1	17676	JPNIC-JP-ASN-BLOCK Japan NIC	86	20448747	27.85%
2	3320	DTAG Deutsche Telekom AG	94	15394896	62.03%
3	3356	LEVEL3 Level 3 Communications	965	9993894	8.60%
4	209	ASN-QWEST - Qwest	2394	5548944	3.65%
5	1668	AOL-ATDN - AOL Transit Data Network	6	5374720	38.55%
6	7018	ATT-INTERNET4 - AT&T WorldNet Services	1568	5080480	2.74%
7	4837	CHINA169-BACKBONE CNCGROUP China169	2062	4484110	21.98%
8	5089	NTL NTL Group Limited	830	4050976	33.92%
9	7132	SBIS-AS - AT&T Internet Services	3849	3403133	11.12%
10	5430	FREENETDE freenet Cityline GmbH	5	3148544	94.20%
11	9121	TTNET TTnet Autonomous System	220	2631676	26.92%
12	12876	AS12876 Telecom Italia France	199	2271707	91.95%
13	6167	CELLCO-PART - Cellco Partnership	6	2163392	7.77%
14	19548	ADELPHIA-AS2 - Road Runner HoldCo LLC	216	2120935	45.83%
15	5617	TPNET Polish Telecom's commercial IP network	723	1983445	50.18%
16	2856	BT-UK-AS BTnet UK Regional network	195	1973216	17.82%
17	3215	AS3215 France Telecom - Orange	907	1862862	18.91%
18	11427	SCRR-11427 - Road Runner HoldCo LLC	408	1846272	60.61%
19	15557	LDCOMNET NEUF CEGETEL (formerly LDCOM)	89	1806588	58.74%
20	3209	Arcor IP-Network	18	1679360	49.57%
21	3269	ASN-IBSNAZ TELECOM ITALIA	288	1667924	14.14%
22	12322	PROXAD AS for Proxad/Free ISP	93	1527036	17.21%
23	11426	SCRR-11426 - Road Runner HoldCo LLC	95	1364992	48.59%
24	4766	KIXS-AS-KR Korea Telecom	1285	1344087	6.23%
25	4812	CHINANET-SH-AP China Telecom (Group)	168	1287165	17.92%
26	6830	UPC UPC Broadband	3671	1280009	21.56%
27	1221	ASN-TELSTRA Telstra Pty Ltd	284	1258748	9.94%
28	3352	TELEFONICA-DATA-ESPANA IAN of TDE	109	1251590	15.63%
29	20001	ROADRUNNER-WEST - Road Runner HoldCo LLC	79	1190240	44.81%
30	6805	TDDE-ASN1 Telefonica Deutschland AS	450	1183279	34.44%
31	3462	HINET Data Communication Business Group	179	1114354	14.41%
32	9105	TISCALI-UK Tiscali UK	42	1110016	45.88%
33	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELES	716	1052005	43.65%
34	12479	UNI2-AS Uni2 Autonomous System	76	992254	28.09%
35	1257	TELE2	447	964642	18.15%
36	5462	CABLEINET Telewest Broadband	346	957690	34.46%
37	11351	RR-NYSREGION-ASN-01 - Road Runner HoldCo LLC	154	933632	21.09%
38	20115	CHARTER-NET-HKY-NC - Charter Communications	816	891873	29.23%
39	4589	EASYNET Easynet Group Plc	125	832841	22.14%
40	7015	CCCH-AS2 - Comcast Cable Communications Holdings,	28	788224	17.32%
41	2510	JPNIC-ASBLOCK-AP JPNIC	788	757694	26.26%
42	852	ASN852 - Telus Advanced Communications	1119	756841	16.53%
43	6739	ONO-AS Cableuropa - ONO	70	745982	48.96%
44	6678	AS-NOOS NOOS Autonomous System	220	727808	34.71%
45	10994	TAMPA2-TWC-5 - Road Runner HoldCo LLC	79	659712	41.78%
46	11530	EMBARQ-MNFD - Embarq Corporation	97	658432	33.46%
47	13343	SCRR-13343 - Road Runner HoldCo LLC	21	654336	54.80%
48	5432	BELGACOM-SKYNET-AS Belgacom regional ASN	142	649719	39.49%
49	174	COGENT Cogent/PSI	859	640697	2.15%
50	12271	SCRR-12271 - Road Runner HoldCo LLC	68	635392	35.69%

**xbl.spamhaus.org**

rank	asid	name	entries	range	quota
1	4837	CHINA169-BACKBONE CNCGROUP China169 Backbon	300498	300498	1.47%
2	9121	TTNET TTnet Autonomous System	227882	227882	2.33%
3	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELES	208809	208809	8.66%
4	8167	TELESC - Telecomunicacoes de Santa Catarina SA	189197	189197	8.83%
5	5617	TPNET Polish Telecom's commercial IP network	188802	188802	4.78%
6	9829	BSNL-NIB National Internet Backbone	132280	132280	6.76%
7	3320	DTAG Deutsche Telekom AG	125390	125390	0.51%
8	3462	HINET Data Communication Business Group	105091	105091	1.36%
9	7643	VNN-AS-AP Vietnam Posts and Telecommunications (V	101830	101830	10.52%
10	4766	KIXS-AS-KR Korea Telecom	97826	97826	0.45%
11	22927	Telefonica de Argentina	88041	88041	11.34%
12	6713	IAM-AS	53256	53256	4.07%
13	3215	AS3215 France Telecom - Orange	48224	48224	0.49%
14	3269	ASN-IBSNAZ TELECOM ITALIA	46895	46895	0.40%
15	3209	Arcor IP-Network	44085	44085	1.30%
16	9737	TOTNET-TH-AS-AP Telephone Organization of Thailand	42411	42411	4.87%
17	4812	CHINANET-SH-AP China Telecom (Group)	41881	41881	0.58%
18	4814	CHINA169-BBN CNCGROUP IP networkÄ;ÄaChina169 B	38192	38192	1.58%
19	4788	TMNET-AS-AP TM Net, Internet Service Provider	36712	36712	0.77%
20	3352	TELEFONICA-DATA-ESPANA Internet Access Network	34033	34033	0.43%
21	18101	RIL-IDC Reliance Infocom Ltd Internet Data Centre,	30082	30082	3.75%
22	8551	BEZEQ-INTERNATIONAL-AS Bezeqint Internet Backbor	29895	29895	2.34%
23	18881	Global Village Telecom	24152	24152	3.97%
24	5483	HTC-AS Hungarian Telecom	23453	23453	3.71%
25	6849	UKRTELNET JSC UKRTELECOM,	23213	23213	18.74%
26	209	ASN-QWEST - Qwest	21419	21419	0.01%
27	8359	COMSTAR COMSTAR-Direct Moscow region network	20680	20680	2.90%
28	9105	TISCALI-UK Tiscali UK	20473	20473	0.85%
29	8452	TEDATA TEDATA	20437	20437	1.90%
30	13184	HANSENET HanseNet Telekommunikation GmbH	20253	20253	1.71%
31	15557	LDCOMNET NEUF CEGETEL (formerly LDCOM NETW	19659	19659	0.64%
32	2856	BT-UK-AS BTnet UK Regional network	19403	19403	0.18%
33	9299	IPG-AS-AP Philippine Long Distance Telephone Compar	19183	19183	2.14%
34	16735	Companhia de Telecomunicacoes do Brasil Central	19147	19147	7.79%
35	4230	Embratel	18383	18383	0.50%
36	17974	TELKOMNET-AS2-AP PT TELEKOMUNIKASI INDONES	18077	18077	2.51%
37	7132	SBIS-AS - AT&T Internet Services	17418	17418	0.06%
38	12876	AS12876 Telecom Italia France	16824	16824	0.68%
39	18403	FPT-AS-AP The Corporation for Financing & Promoting T	16692	16692	5.43%
40	9304	HUTCHISON-AS-AP Hutchison Global Communications	16191	16191	0.53%
41	15475	NOL	16164	16164	2.32%
42	4808	CHINA169-BJ CNCGROUP IP network China169 Beijing	16021	16021	0.32%
43	5391	T-HT T-Com Croatia Internet network	16014	16014	3.24%
44	9394	CRNET CHINA RAILWAY Internet(CRNET)	14963	14963	0.12%
45	19429	ETB - Colombia	14810	14810	4.37%
46	5486	SMILE-ASN Euronet Digital Communications, (1992) LTD	14667	14667	1.11%
47	7552	VIETEL-AS-AP Vietel Corporation	14635	14635	0.34%
48	17858	KRNIC-ASBLOCK-AP KRNIC	14284	14284	0.32%
49	9050	RTD RTD-ROMTELECOM Autonomous System Number	14266	14266	3.15%
50	20858	EGYNET-AS	14002	14002	8.48%

**dnswl.org**

rank	asid	name	entries	range	quota
1	559	SWITCH SWITCH, Swiss Education and Research Netw	47	197162	8,09%
2	3303	SWISSCOM Swisscom Solutions Ltd	108	139431	3,72%
3	3598	MICROSOFT-CORP-AS - Microsoft Corp	3	66048	9,09%
4	702	AS702 Verizon Business EMEA - Commercial IP service	51	65886	0,26%
5	786	JANET The JANET IP Service	40	65575	0,89%
6	721	DISA-ASNBLK - DoD Network Information Center	30	65565	0,07%
7	12257	DGC - Data General Corporation	3	65538	80,51%
8	4583	WESTPUB-A - West Publishing Corporation	2	65537	45,88%
9	12701	Barclays Capital Autonomous System	1	65536	72,73%
10	16780	Banco SantanderSantiago	1	65536	50,00%
11	16729	AS16729 - Royal Bank of Canada	1	65536	71,31%
12	15675	ETAT-DE-VAUD Etat de Vaud, CCT	1	65536	94,47%
13	25215	BNP-PARIBAS BNP PARIBAS	1	65536	96,97%
14	20617	BNP-PARIBAS AS for BNP Paribas UK Ltd	1	65536	33,38%
15	25180	EXPONENTIAL-E-AS Exponential-e Ltd	1	65536	58,45%
16	7734	TDBANK - Toronto Dominion Bank	1	65536	91,10%
17	8075	MICROSOFT-CORP---MSN-AS-BLOCK - Microsoft Corp	31	16906	3,89%
18	14779	INKTOMI-LAWSON - Inktomi Corporation	15	10246	25,17%
19	15576	NTS NTS workspace AG, Bern, Switzerland	3	8194	34,05%
20	9732	SCIG-AS-AP CENTRAL INTERNET SERVICES	1	8192	47,76%
21	13267	ZKB Zuercher Kantonalbank	1	8192	100,00%
22	29500	SWISSRE-AS Schweizerische Rueckversicherungsgese	2	8192	100,00%
23	14780	INKTOMI-LAWSON - Inktomi Corporation	22	5377	18,59%
24	15635	AS15635 Yahoo! Europe AS	4	4864	51,35%
25	1273	CW Cable & Wireless	14	4410	0,09%
26	33845	SWISSGOV Swiss Government	6	4356	3,18%
27	10389	BEARCLEARNET - Bear Stearns Security Corporation	1	4096	72,73%
28	21069	ASN-METANET METANET AG, Switzerland	32	3857	47,08%
29	1668	AOL-ATDN - AOL Transit Data Network	26	3149	0,02%
30	26101	YAHOO-3 - Yahoo!	12	3072	18,18%
31	6730	SUNRISE sunrise (TDC Switzerland AG)	32	2967	0,61%
32	3561	SAVVIS - Savvis	203	2398	0,03%
33	7018	ATT-INTERNET4 - AT&T WorldNet Services	146	2310	0,00%
34	8068	MICROSOFTEU Microsoft European Data Center	4	2051	15,41%
35	36017	SIXAPART - SIX APART LTD	1	2048	100,00%
36	12903	GARANTI-TECH Garanti Bank, Turkey	1	2048	32,00%
37	2134	GSVNET-AS GS Virtual Network	1	2048	20,00%
38	8560	ONEANDONE-AS 1&1 Internet AG	20	1551	0,45%
39	5432	BELGACOM-SKYNET-AS Belgacom regional ASN	7	1537	0,09%
40	15625	ING-AS ING NV (ITC)	3	1536	2,34%
41	10361	BLOOMBERG-NET - Bloomberg, LP	5	1536	11,54%
42	13030	INIT7 Init Seven AG, Zurich, Switzerland	42	1331	0,48%
43	15623	CYBERLINK Cyberlink Internet Services AG	40	1322	1,75%
44	10228	YAHOO-CN Internet Content Provider	5	1280	16,67%
45	21345	MESSAGELABS Messagelabs Anti Virus Solutions	3	1280	83,33%
46	17110	YAHOO-US2 - Yahoo	5	1280	11,36%
47	29097	HOSTPOINT-AS Hostpoint AG, Switzerland	5	1280	31,25%
48	3356	LEVEL3 Level 3 Communications	70	1089	0,00%
49	1239	SPRINTLINK - Sprint	39	1060	0,00%
50	36752	YAHOO-SP1 - Yahoo	19	1039	4,36%

**Union of all blacklists**

rank	asid	name	entries	range	quota
1	17676	JPNIC-JP-ASN-BLOCK Japan Network Information Cent	217	20448878	27.85%
2	3320	DTAG Deutsche Telekom AG	893	15395695	62.04%
3	3356	LEVEL3 Level 3 Communications	4952	9997881	8.60%
4	209	ASN-QWEST - Qwest	7128	5553678	3.65%
5	1668	AOL-ATDN - AOL Transit Data Network	14	5374728	38.55%
6	7018	ATT-INTERNET4 - AT&T WorldNet Services	8812	5137125	2.77%
7	4837	CHINA169-BACKBONE CNCGROUP China169 Backbon	79476	4561014	22.36%
8	5089	NTL NTL Group Limited	3591	4053737	33.94%
9	7132	SBIS-AS - AT&T Internet Services	39005	3438800	11.24%
10	5430	FREENETDE freenet Cityline GmbH	815	3149354	94.22%
11	9121	TTNET TTnet Autonomous System	6957	2638668	26.99%
12	12876	AS12876 Telecom Italia France	465	2271973	91.96%
13	6167	CELLCO-PART - Cellco Partnership	1244	2164630	7.78%
14	19548	ADELPHIA-AS2 - Road Runner HoldCo LLC	2206	2122925	45.87%
15	2856	BT-UK-AS BTnet UK Regional network	19755	1992776	18.00%
16	5617	TPNET Polish Telecom's commercial IP network	6139	1988861	50.32%
17	3215	AS3215 France Telecom - Orange	26415	1888370	19.17%
18	11427	SCRR-11427 - Road Runner HoldCo LLC	1365	1847229	60.64%
19	15557	LDCOMNET NEUF CEGETEL (formerly LDCOM NETW	889	1807388	58.77%
20	3209	Arcor IP-Network	4039	1683381	49.69%
21	3269	ASN-IBSNAZ TELECOM ITALIA	7154	1675045	14.20%
22	12322	PROXAD AS for Proxad/Free ISP	6726	1533669	17.29%
23	4766	KIXS-AS-KR Korea Telecom	164003	1506805	6.98%
24	11426	SCRR-11426 - Road Runner HoldCo LLC	845	1365742	48.62%
25	4812	CHINANET-SH-AP China Telecom (Group)	9635	1296632	18.05%
26	6830	UPC UPC Broadband	7598	1283689	21.62%
27	3352	TELEFONICA-DATA-ESPANA Internet Access Network	19653	1271389	15.88%
28	1221	ASN-TELSTRA Telstra Pty Ltd	7222	1265686	9.99%
29	20001	ROADRUNNER-WEST - Road Runner HoldCo LLC	383	1190544	44.82%
30	6805	TDDE-ASN1 Telefonica Deutschland Autonomous Syste	1113	1183942	34.45%
31	3462	HINET Data Communication Business Group	4918	1119093	14.47%
32	9105	TISCALI-UK Tiscali UK	1455	1111429	45.94%
33	27699	TELECOMUNICACOES DE SAO PAULO S/A - TELES	7282	1058571	43.92%
34	12479	UNI2-AS Uni2 Autonomous System	207	991874	28.07%
35	1257	TELE2	6308	970503	18.26%
36	5462	CABLEINET Telewest Broadband	5058	962402	34.63%
37	11351	RR-NYSREGION-ASN-01 - Road Runner HoldCo LLC	705	934183	21.10%
38	20115	CHARTER-NET-HKY-NC - Charter Communications	8781	899838	29.49%
39	4589	EASYNET Easynet Group Plc	1101	833817	22.16%
40	7015	CCCH-AS2 - Comcast Cable Communications Holdings,	119	788315	17.33%
41	2510	JPNIC-ASBLOCK-AP JPNIC	3353	760259	26.35%
42	852	ASN852 - Telus Advanced Communications	3417	759139	16.58%
43	6739	ONO-AS Cableuropa - ONO	962	746874	49.02%
44	6678	AS-NOOS NOOS Autonomous System	802	728390	34.73%
45	577	BACOM - Bell Canada	10899	711282	16.51%
46	10994	TAMPA2-TWC-5 - Road Runner HoldCo LLC	703	660336	41.82%
47	11530	EMBARQ-MNFD - Embarq Corporation	354	658689	33.47%
48	13343	SCRR-13343 - Road Runner HoldCo LLC	852	655167	54.87%
49	5432	BELGACOM-SKYNET-AS Belgacom regional ASN	1504	651081	39.57%
50	174	COGENT Cogent/PSI	1687	641525	2.16%

## F Annex - Graphical blacklist coverage

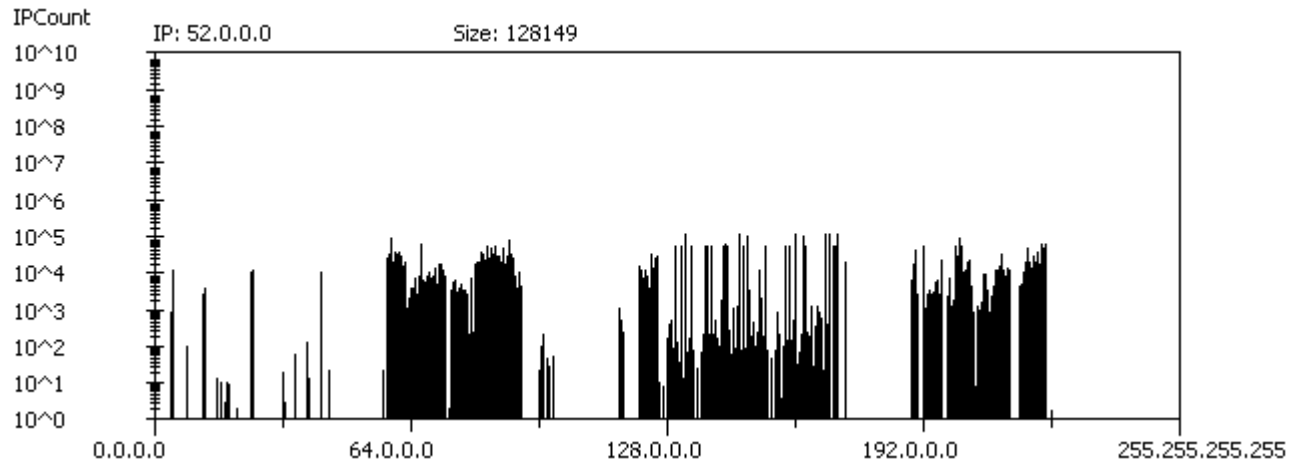


Figure 23: all.dnsbl.sorbs.net

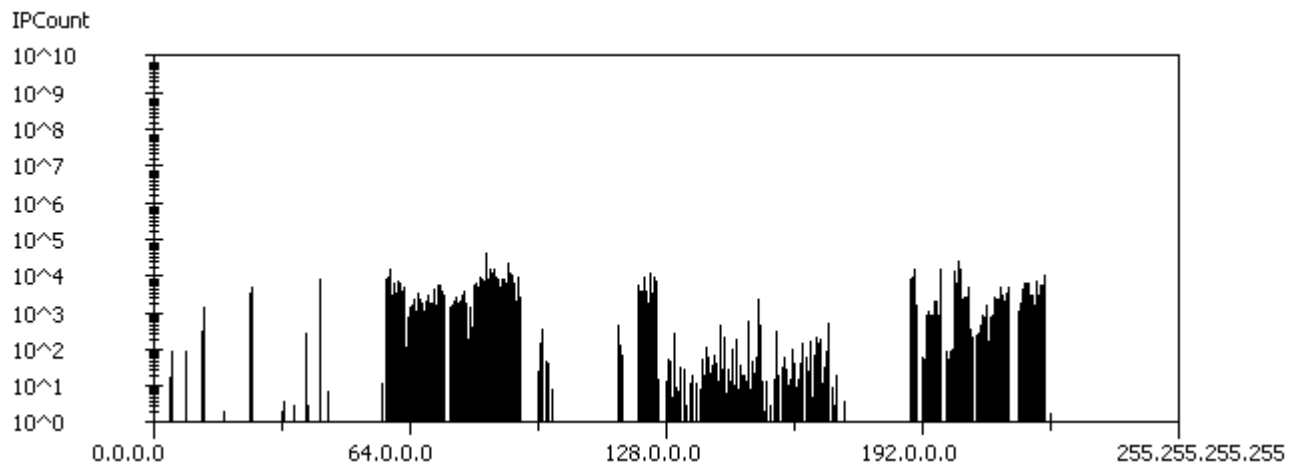


Figure 24: UCEPROTECT - Level 1

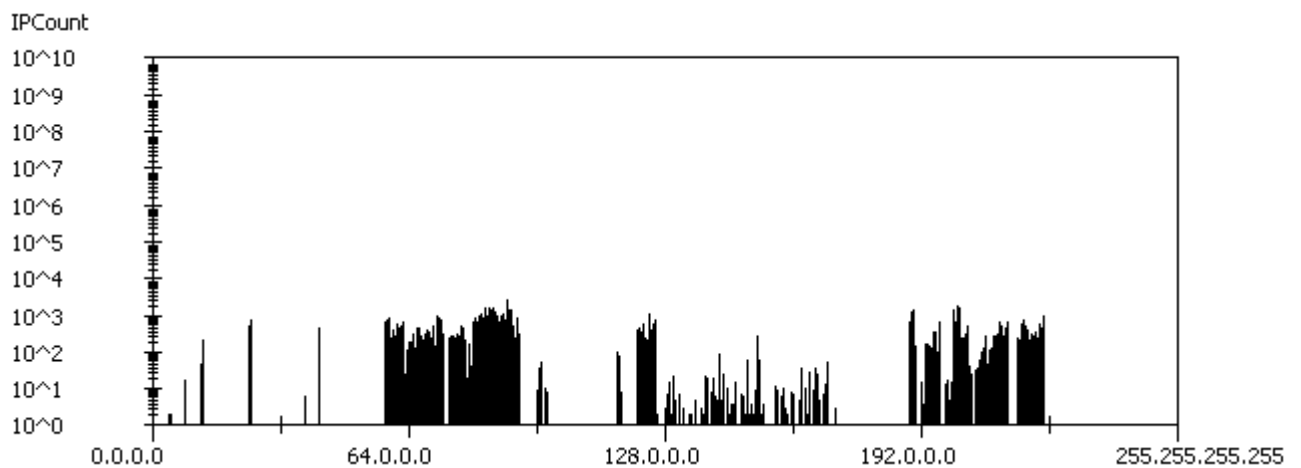


Figure 25: NiX Spam

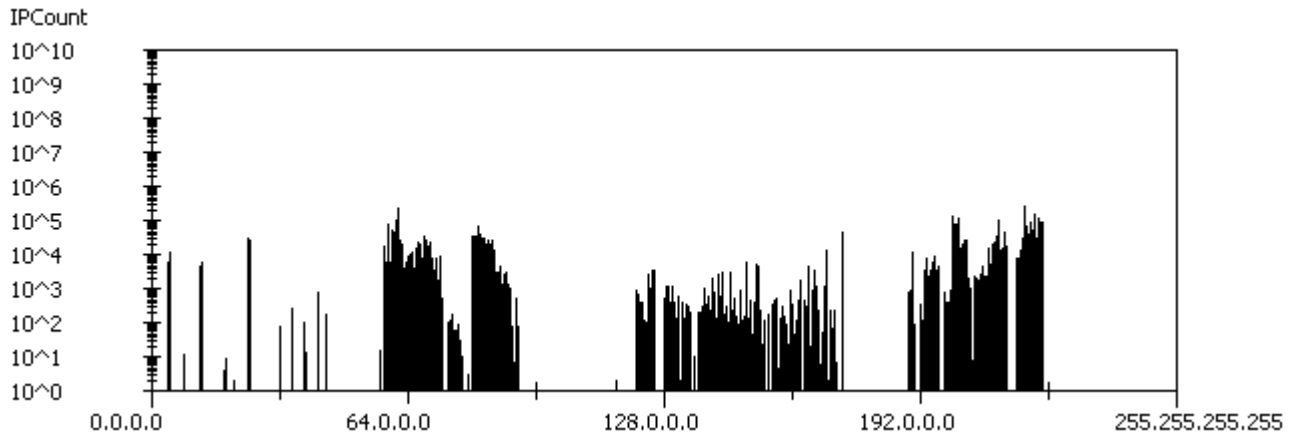


Figure 26: dnsbl.ahbl.org

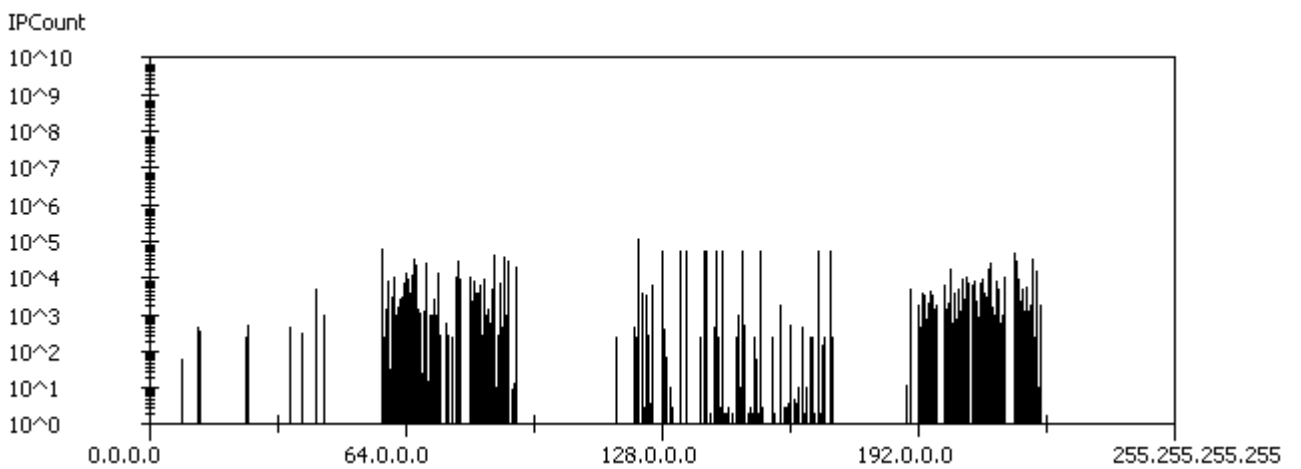


Figure 27: sbl.spamhaus.org

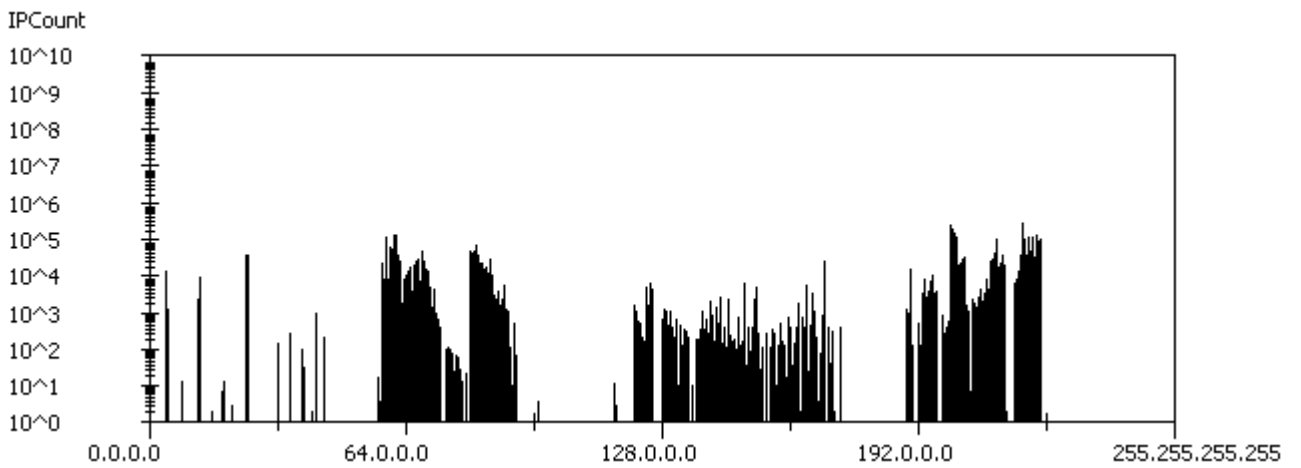


Figure 28: dnsbl.njabl.org

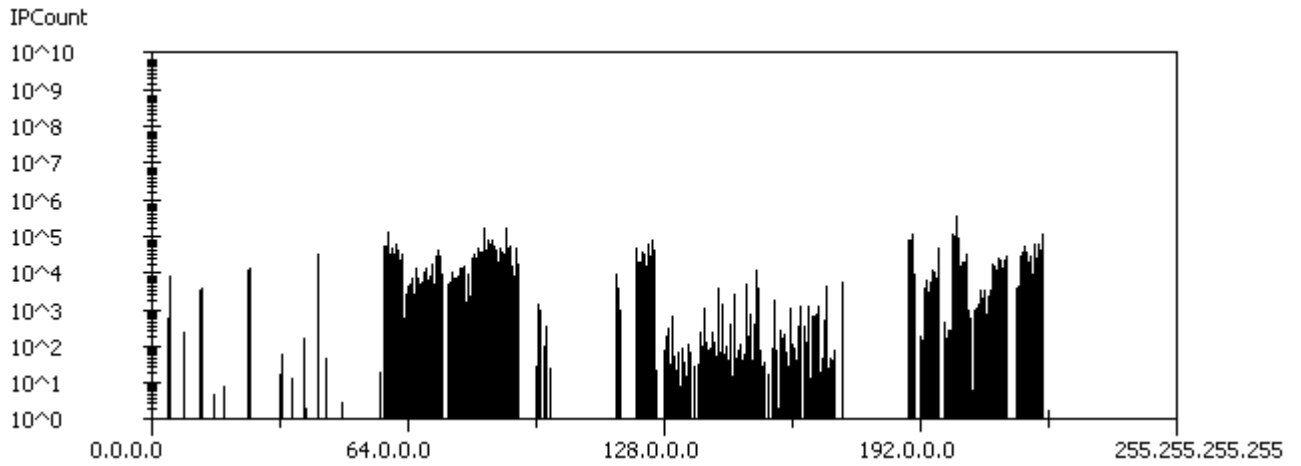


Figure 29: CBL

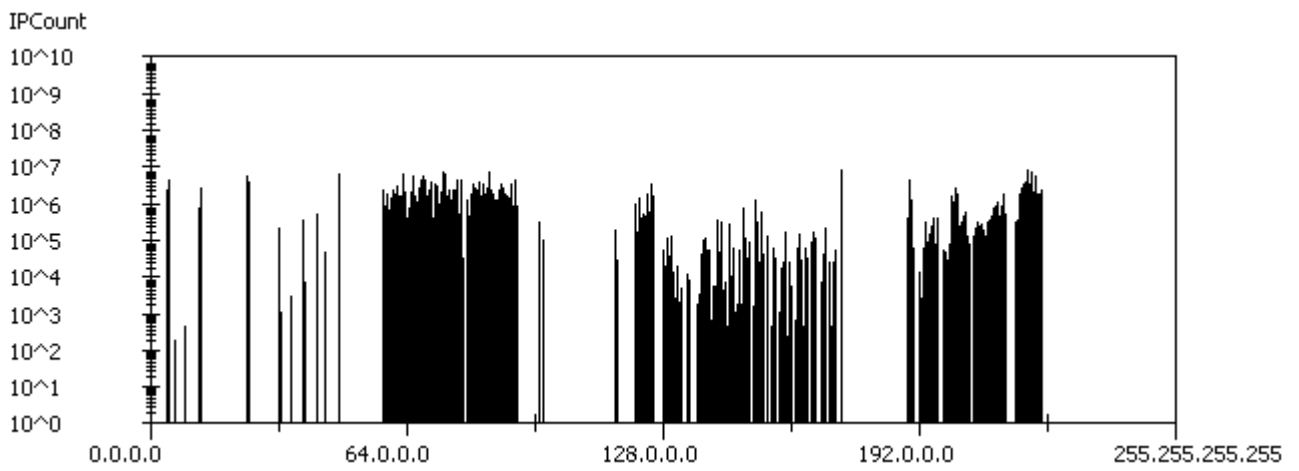


Figure 30: pbl.spamhaus.org

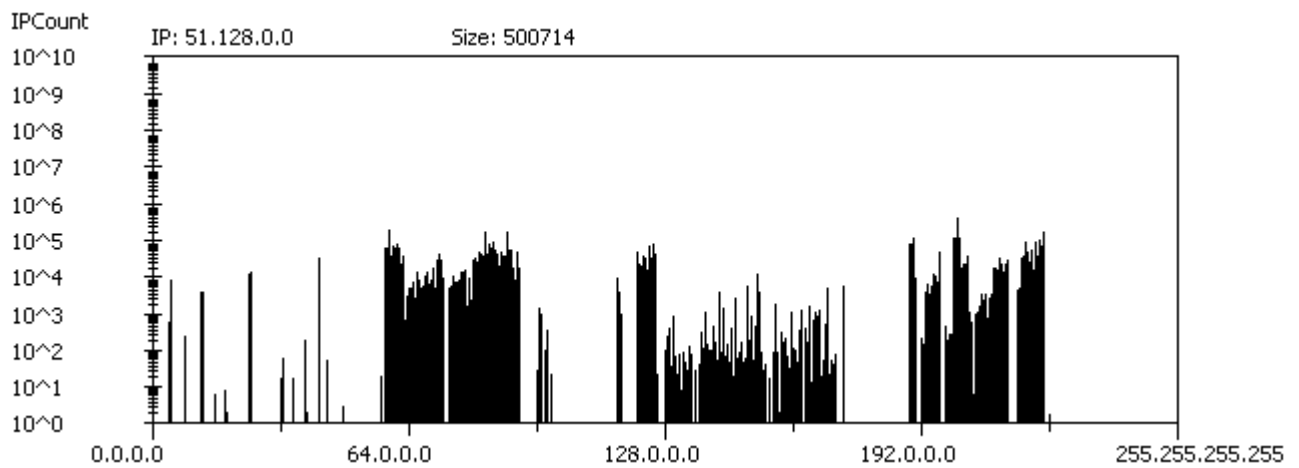


Figure 31: xbl.spamhaus.org