

E-mail Spam – 30 Years After

A SHORT (HISTORIC) OVERVIEW



Disclaimer

The information and data asserted in this document represent the current opinion of BitDefender® on the topics addressed as of the date of publication. This document and the information contained herein should not be interpreted in any way as a BitDefender's commitment or agreement of any kind.

Although every precaution has been taken in the preparation of this document, the publisher, authors and contributors assume no responsibility for errors and/or omissions. Nor is any liability assumed for damages resulting from the use of the information contained herein. In addition, the information in this document is subject to change without prior notice. BitDefender, the publisher, authors and contributors cannot guarantee further related document issuance or any possible post-release information.

This document and the data contained herein are for information purposes only. BitDefender, the publisher, authors and contributors make no warranties, express, implied, or statutory, as to the information stated in this document.

The document content may not be suitable for every situation. If professional assistance is required, the services of a competent professional person should be sought. Neither BitDefender, the document publishers, authors nor the contributors shall be liable for damages arising here from.

The fact that an individual or organization, an individual or collective work, including printed materials, electronic documents, websites, etc., are referred in this document as a citation and/or source of current or further information does not imply that BitDefender, the document publisher, authors or contributors endorses the information or recommendations the individual, organization, independent or collective work, including printed materials, electronic documents, websites, etc. may provide. Readers should also be aware that BitDefender, the document publisher, authors or contributors cannot guarantee the accuracy of any information presented herein after the date of publication, including, but not limited to World Wide Web addresses and Internet links listed in this document which may have changed or disappeared between the time this work was written and released and the moment it is read.

The readers are entirely responsible to comply with all applicable international copyright laws arising from this document. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of BitDefender.

BitDefender may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from BitDefender, this document does not provide any license to these patents, trademarks, copyrights, or other intellectual property.

Copyright © 2008 BitDefender. All rights reserved.

All other product and company names mentioned herein are for identification purposes only and are the property of, and may be trademarks of, their respective owners.

E-mail Spam – 30 Years After

Răzvan Livintz
Communication Specialist

E-mail spam represents an unsolicited message, usually (but not always) of a commercial nature, indiscriminately sent to you and to a large number of other recipients by an unknown sender.

If you are reading these lines, you are probably one of the 1,407,724,920 people from this planet, according to a statistic published couple months ago¹, who own a computer and get online quite often. Chances are that you received this article or a link directing you here within an e-mail. And, most likely, this should also be true for the other several readers laying their eyes on the same paragraph by now. You probably noticed some of them, by their names and e-mail addresses appearing next to yours in the header of the message I just mentioned. And, almost certainly, at least one or two guys from the *To* field are total strangers for you, if not all of them (not to mention the name appearing in the *From* field which, by the way, is not mine). There is also a very good chance for you not to be interested at all in receiving and reading this article, in which case I must humbly apologize in advance for wasting your time and patience, and respectfully assure you that this was not in any way my intention.

Technically speaking, an e-mail message bearing the features I outlined in the previous paragraph is called *spam* or *bulk e-mail* or just *junk e-mail*. An utterly unpleasant (post-)modern times invention that you, and me, and one fifth of the (Internet connected) world “enjoy” each morning in our daily routine, when we open our e-mail inbox.

Why do we call it spam?

Monty Python’s sketch revolves around a lot of...
spam:

Waitress: Well, there’s egg and bacon; egg sausage and bacon; egg and spam; egg bacon and spam; egg bacon sausage and spam; spam bacon sausage and spam; spam egg spam spam bacon and spam; spam sausage spam spam bacon spam tomato and spam...

If you are willing to make an experiment and “google™” the word *spam*, the first result the search engine displays is the website of SPAM® luncheon meat from Hormel® Foods Corporation. Apparently, the unsolicited e-mail avalanches have nothing to do with the canned pork meat.

However, the alleged link between the two is to be found in a popular culture reference of the 70s, to be more specific, in a three and a half minutes sketch of Monty Python’s British television show. Obviously, the sketch name is *Spam*² and it depicts the troubles of a couple who is trying to order breakfast from a menu overwhelmed with all sort of dishes including spam.

A quarter century later, the discussion system developed in the early 80s at University of North Carolina at Chapel Hill and Duke University, USENET, the ancestor of our World Wide Web, got flooded by two immigration lawyers from Phoenix, Laurence Canter and Martha Siegel³, his wife. They posted the same message advertising their services in an upcoming green card lottery on thousands of newsgroups, pretty much the same way the characters from Monty Python’s sketch incessantly repeated the word *spam*.

¹ As asserted by the *Internet World Stats*, last update 28 May 2008, Miniwatts Marketing Group, accessed last time 30 May 2008, <http://www.internetworldstats.com/stats.htm>.

² As you probably expected, the sketch is available on YouTube: Monty Python, *Spam*, added by zumpzump 14 February 2007, *YouTube*, accessed last time 06 June 2008, <http://www.youtube.com/watch?v=anwy2MPT5RE>. Do not skip the scrolling end credits!

³ For detailed media coverage of their “enterprise”, see Canter’s own website, especially Laurence Canter, “Profile of Laurence A. Canter”, last update 31 May 2005, *Web Site of Laurence A. Canter*, accessed last time 06 June 2008, <http://www.l-ware.com/press.htm>.

Although this is not the first time when an event like this occurred, it is however the first time⁴ when someone thought to associate the word *spam* to this kind of action and its consequences.

When did spam appear?

I am quite sure you have noticed in the introduction the prefix *post-* placed between parentheses. This is because the first spam attempts date most likely quite long before the age of Internet, no matter how strange this would look like.

If we agree that spam is always unsolicited and targets quite a lot of recipients, I think we can probably qualify as spam the first leaflets ever printed and dropped in numerous “classic” postal mail boxes (an “old” nasty habit we still experience today, but under the name of “printed advertising materials”).

Although postal systems (and, subsequently, mail boxes) appeared around the second half of the 3rd millennium B.C. in Egypt, and Gutenberg’s printing machine emerged four thousand years later, in the second half of the 15th century, we might trace back different forms of spam in the early mankind days: from the Egyptian advertising papyruses or the Ancient Greece and Rome graffiti and incised pottery, to the printed selling announces, ballads and political pamphlets of the 17th and 18th century fliers.

With the establishment of the national and regional post services in the early and mid 19th century, commercial handbills type of spam entered a new age, being increasingly exploited later on, during World Wars, but not limited to, as a form of airborne propaganda through the so-called leaflet bombs.

The egression of new communication media also carried along the spam convolutions, whether we talk about the telegraph, fax, phone or mobile phone services, and last but not least e-mail.

Even if the e-mail spam seems to be “the youngest” type of spam, it is quite old. 30 years old, actually! The *first e-mail spam* is the “son” of marketing manager Gary Thuerk and engineer Carl Gartley, from East Coast-based Digital Equipment Corporation. Thuerk, the first known e-mail spammer, wanted to advertise throughout the Defense Advanced Research Projects Agency Network (ARPANET) West Coast members the presentations his company held for several new computer models. His message was sent to almost 400 ARPANET users of the 2600 people who had an e-mail account in those days.

Thuerk does not believe he was the first spammer: “Actually, I think of myself as the father of e-marketing. There’s a difference. (...) E-spam is a blast of unsolicited e-mail and/or malware to an unqualified list of recipients. It is unwanted by almost all of those who receive it,” says Thuerk. As opposed to e-mail spam, e-marketing targets several people who “have a known or qualified interest in your product, service or the information you are sending”, he adds⁵.

In contrast to current days’ trend, Thuerk’s mail was quite a success, his initiative bringing to Digital Equipment Corporation several million dollars from sales based on that message.

⁴ For other theories concerning the etymology of the word *spam*, please see the excellent article of Brad Templeton, “Origin of the term «spam» to mean net abuse”, *Brad Templeton’s Home Page*, accessed last time 30 May 2008, <http://www.templetons.com/brad/spamterm.html>, as well as his other essays available in the same location.

⁵ Gina Smith, “Unsung innovators: Gary Thuerk, the father of spam”, published 03 December 2007, in *Computerworld*, accessed last time 06 June 2008, <http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9046419>.

Why do we get spam anyway?

*Top 10 of 2008's most
advocated content
through e-mail spam*

01. Drugs
02. Replica Watches
03. Phishing (tool for)
04. Pirated Software
05. Pornography
06. Loans
07. Hire & Employment
08. Trojan Viruses Spread
(tool for)
09. Dating
10. Diploma

*Source: BitDefender
Antispam Lab*

From May, 3rd, 1978, when Thuerk sent his message until today, e-mail spam's main reason is pretty much the same. PROFIT!

Whether we talk about advertising special goods, announcing huge discounts on all different kind of drugs and their substitutes, promoting new social networks or last-minute hard-core porn websites, and so on, all spam creators and disseminators seek to gain some profit – either by determining you to buy the advertised goods and services, either by wasting (and sometimes neutralizing) your resources.

It is something we also do on a regular basis (probably without being aware of it) when sending a message to anybody from our address books like Thuerk did 30 years ago. Probably not all our friends are really interested about the stuff we “advertise” (not to mention chain letters, fake raising funds petitions for earthquakes victims, jokes and other similar “rubbish”). We do it for our profit – just to save some time (or to be fortunated in the future).

How does spam get in our inboxes?

In the early days of e-mail spam, the unsolicited messages were sent out directly, to the addresses the spammer grabbed. Eventually, the address or the IP he or she used was tracked back and blocked. This led to address and domain spoofing, as well as counterfeiting other “honest” e-mail elements to deceive the vigilance of e-mail server administrators.

Later on, the spam distribution mechanisms switched from the open relay e-mail servers (where a sender could, technically, mail any recipient) and modem pools (where spammers exploited ISP dial-up services' vulnerabilities) to the cable and ADSL modem based proxy servers.

Contemporary distribution methods are even more sophisticated, more aggressive, and, most important, automated.

The essential element for spamming is not, in fact, the spam message itself, but the *list of e-mail addresses*. Without a large number of recipients, the message content is worthless and useless. 21st century spammers no longer “dig” for e-mails in white/yellow pages or directories, like 30 years ago, but employ automated tools, such as *crawlers*, *spiders*⁶ and *bots*⁷. To be more accurate *spambots*⁸. The malicious job of all these automated tools is to “harvest” e-mail addresses from Web pages, contact forms, guestbook pages, mailing lists, shopping and gift lists, as well as other sources which may contain this type of data, including (but not limited to) our business and personal address books like those we store in an e-mail client such as Microsoft® Outlook®, Mozilla® Thunderbird™, Qualcomm® Eudora®, etc.

This is why, technically speaking, the cybercriminals involved in spam activities may be divided in two major categories, based on the nature of their occupa-

“Most of the harvesting and spamming-related activities can be done from any corner of the world, as long as you have a computer and an Internet connection. This is why, tracking and apprehending a harvester or spammer is like looking for a needle in a haystack.”

Claudiu Muşat

*BitDefender
Antispam Researcher*

⁶ (*Web*) *crawler* and (*Web*) *spider* refer to an application or automated script which surfs the World Wide Web and gathers or collects specific information. The benign use of crawlers concerns the search engines' data feeding process, understood as basis for further indexing operations. The malign use is, of course, related to information's illicit retrieval and usage.

⁷ *Bot* represents an abbreviation of *robot* or *Internet robot*, designating a software application that can perform several automated and repetitive tasks, usually over the Internet.

⁸ *Spambot* is a coined term from *spam* and the abbreviated *robot* and refers to a program designed to scan and fetch hyperlinks and e-mail addresses from Web pages, usually for further use. Sometimes they are just a small part of a larger application which combines automated information retrieval and additional data transmission, such as e-mail spam distribution.

tion. On one hand, we have the *harvesters* – those people responsible for the e-mail addresses' collection. On the other hand, we find the *spammers* – those to blame for the unsolicited messages distribution. Frequently, these two roles belong to the same person, although there are also individuals who limit their illicit actions to harvesting and underground trade only, because they consider it more lucrative than the spam delivery itself.

Try to imagine what a gold mine for a harvester would be the customers' e-mail list of a major online shop, with millions of active and valid addresses (not to mention other sensitive personal information, such as credit card numbers, for instance), multiplied with a certain amount of money per address.

“Many forums, discussion groups, blogs or guest books run on Web-based applications. When an exploit affects all these media, a harvester can extract with no trouble at all an incredible amount of e-mail addresses from thousands and thousands of compromised Web pages. In addition, if the vulnerability is not publicly disclosed (so that the users take the necessary measures), and the application producer does not release a security patch, the harvester may take advantage of it for a longer timeframe.”

George Petre

BitDefender Spam Intelligence Researcher

Another method widely spread for e-mail addresses' harvesting is to exploit Web sites vulnerabilities. Different portals provide their visitors free access accounts, as well as subscriptions to periodic newsletters, bulletins, announcements, etc. Without the appropriate security layer to defend their content, these Web sites are an easy pray for any harvester. By using a method such as SQL injection⁹, a (not too skilled) harvester can obtain without too much difficulty a brand new list, with thousands of perfectly valid addresses, almost 100% proved to be active.

But not only the World Wide Web and the content posted through its pages represent a source for e-mail addresses. Although sometimes it is a supply for malware and spam by itself¹⁰, file sharing via peer-to-peer networks¹¹ is often a simple and reduced cost solution for a harvester looking for e-mail addresses. The reason is very simple – many P2P users accidentally share their root folders and subfolders (such as the entire C:\ drive, including Windows and Program Files). Most often, these folders hold the e-mail client's address book: Microsoft® Outlook®, for instance, stores by default the address book under the following path `C:\Documents and Settings\[user]\Application Data\Microsoft\Address Book\`.

In addition to spambots exploitation and other methods, contemporary harvesters and spammers may use some automated scripts that can generate names' and domains' combinations which are probably to appear in an e-mail address.

⁹ *SQL (Structured Query Language) Injection* refers to a technique or type of malicious attack which exploits security vulnerabilities within an application's database layer. It allows an attacker to add or “inject” unexpected SQL code to a Web form and thus to gain access to and manipulate the database content.

For instance, any on-line shop displays a login Web page for customers' authentication. The alphanumeric strings you insert as the username and password in the corresponding form fields are included in a query which addresses the database. If the combination between your username and password matches the one stored within the database, you gain further access and you are allowed to start shopping, place orders and manage other account settings. On the other hand, if the username and password pair you insert does not match any of the possible combinations from the database, your access is denied.

The scenario I just outlined represents the normal usage of the Web form, where the database expects from the user those two specific alphanumeric strings as part of the SQL query addressing the database. The unexpected code previously mentioned, the one a harvester or other cyber-criminal would employ to gain database access, refers to specific SQL instructions he or she might input in the Web form fields instead of the username and password. Although the database rejects an incorrect combination, it might just accept and process one or several SQL statements, which are, in effect, alphanumeric strings. This happens because the Web forms and databases behind them cannot block or reject by themselves such input (i. e. other than username and password). Thus, the harvester can easily get access to the database content and retrieve or alter the information stored within.

¹⁰ If we think to the almost unlimited number of messages the administrative bots send through out the pop-up windows invading the screen.

¹¹ *Peer-to-peer* (abbreviated *P2P*) networks refer to a special decentralized type of network where equal peer computers use a common application to connect with each other and function both as clients and servers for the other machines within the network, providing thus direct access to the shared files and folders. Although this definition sounds very technical, I am pretty sure that you at least heard about the names of the most popular P2P clients, such as Kazaa, eMule or DC++.

Think about your organization's e-mail policy, which most likely combines your first name, surname and/or their initials.

Last but not least, e-mail addresses can be bought. Harvesters and spammers are all social individuals, and like everybody else do love to shop. E-mail lists can be either traded or purchased on the underground market, in exchange for other address lists, credit card numbers, pirated software, or money.

Once the addresses are collected, the spam offensive can be launched. Since the dissemination medium is the e-mail message, the spammers would need an *e-mail account*. One or several. The easiest way to get one is, as we all know, to use a free web-based e-mail service, such as Yahoo!® or Hotmail®. But since they need to send out huge amounts of messages, chances are to use multiple accounts. Here again, the automated tools – bots – come in hand.

But bots can serve other spamming purposes as well. Like gaining access on other people's computers and using them as platforms for sending the e-mail messages. For instance, through a worm, Trojan horse, other virus types or any security breach, the spammer places a bot software inside the machine of an innocent user¹². The bot software usually contains *spreaders* that automate the task of vulnerability scanning. Once found, the unprotected computers are attacked and infected, starting a new bot distribution cycle. A collection of malicious bots whose purpose is to run different kind of computer applications controlled by the owner or the disseminator of the software robot source, on a group of compromised computers, usually connected to the Internet, is referred as a *botnet*¹³. The robot-controlled machines are known as *zombies* or *drones*, while the botnet owner is referred to as *herder*. A botnet may be small or large, depending on the complexity and sophistication of the bots the herder employs. A large botnet consists of thousands individual zombies. A small botnet counts only a several hundred drones. Multiply these figures with those of addresses we store within our address books and with a specific amount of spam messages per day and you can get an image of the quantity of e-mails¹⁴ a spammer may send without too much effort or costs¹⁵.

What does spam look like today?

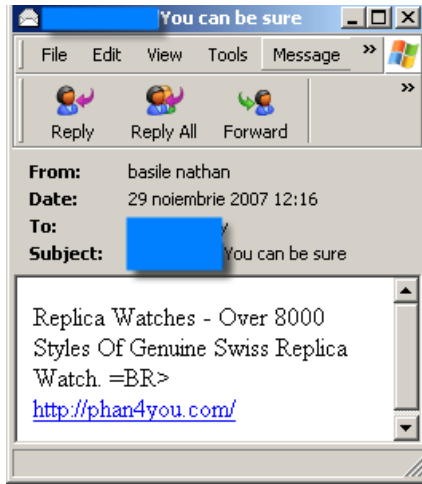
In a certain way, most of the 21st century's spam messages have the early days' characteristics. Meaning that they are mainly either "plain vanilla" *text* or simple HTML text messages, promoting different goods and services like those from the images below:

¹² Although the (in)famous *Win32.Sobig.F@mm*, accountable for multi-billion dollars losses, is recognized as a conjunction or hybrid between a worm and a Trojan, I think that, based on the complexity of distribution mechanism and its damaging capabilities, it might be consider as a precursor of the present-day bots. For a detailed description, please see "Virus Information for - Win32.Sobig.F@mm", in *BitDefender's Virus Encyclopedia*, accessed last time 09 June 2008, <http://www.bitdefender.com/VIRUS-1519-en--Win32.Sobig.F@mm.html>.

¹³ *Botnet* is another coined term derived from *robot network*.

¹⁴ For an example of spam abuse via botnet, see "BitDefender Captures Spam Used to Influence U.S. Presidential Election", published 29 October 2007, in *BitDefender*, accessed last time 09 June 2008, <http://news.bitdefender.com/NW607-en--BitDefender-Captures-Spam-Used-to-Influence-U.S.-Presidential-Election.html>.

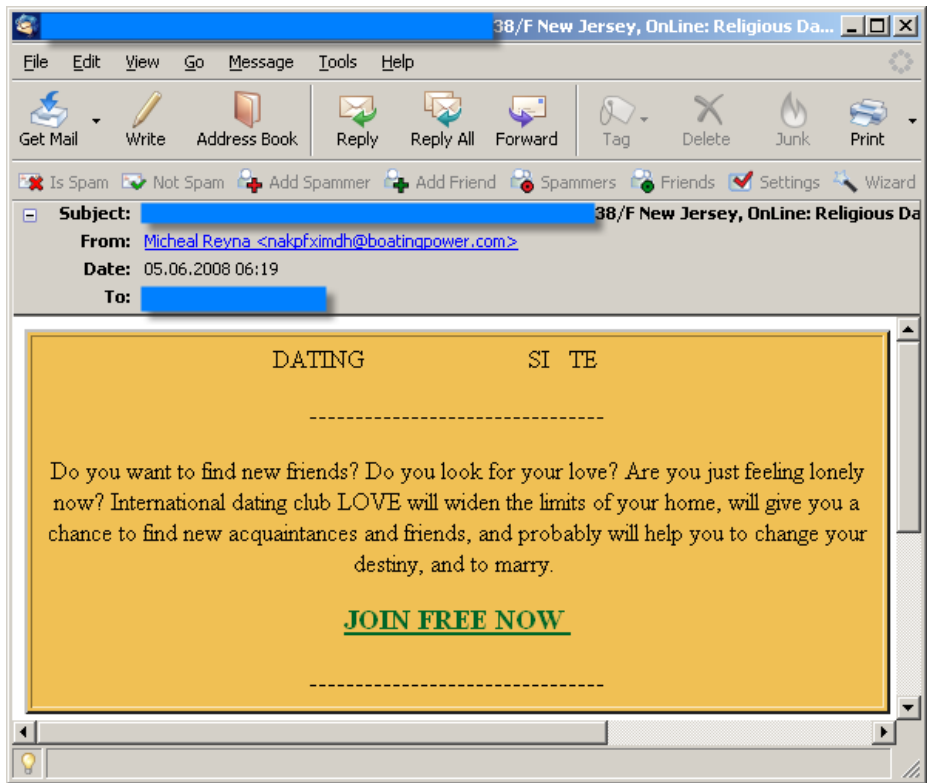
¹⁵ Actually, the effort belongs almost entirely to the exploited drones, since most of their resources (like memory, processing speed, Internet bandwidth, etc.) are drained out by the bots' mischievous actions. As for the costs, I believe it is obvious that they inflict the victim user(s).



“Text-based spam is the most prolific type of the spam family, due to its simplicity, reduced size, and extreme versatility.”

Andra Miloiu

*BitDefender
Spam Analyst*



The message body varies from few words to several phrases or paragraphs and includes often a link directing to the Web site selling the products or hosting the announced services.

Because spam filters can be taught to block messages containing specific words most likely to appear in the subject line or body, one of the techniques that spammers employ is to alter, misspell, replace or add word characters. Although the spam filters cannot detect some of them, because of their unusual pattern, we, as humans, can easily recognize the encrypted words, as displayed in the following image:

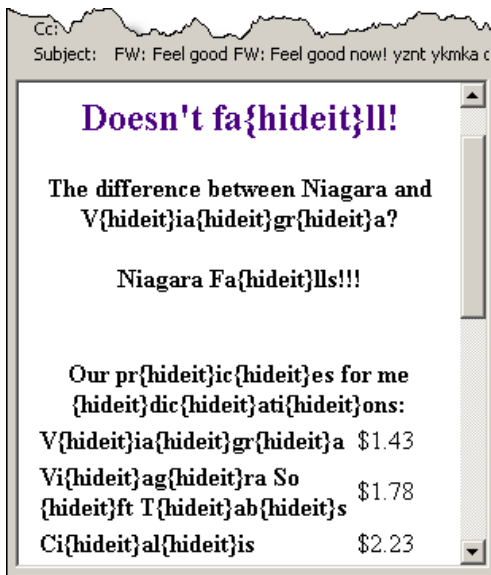


To deceive the content-based spam filters¹⁶, sometimes the same message body and its subject line are combined, partially altered, and switched, via automated scripts.

“If we consider Anatrims attacks (entire waves of spam advertising diet pills back in 2007), that used just 40 samples of different subjects and 3,000 rephrases of the e-mail title (the first centered paragraph), the spammer easily obtained up to 120,000 individual combinations, using only two features as a variation method. Imagine what happens when the body also varies.”

Cătălin Coșoi

BitDefender
Antispam Researcher



This produces a virtual infinite range of unique messages, rather than a single, pattern-based message with a worldwide distribution. For instance, let’s assume a spammer has several text sources (even legitimate ones), as pools for the message’s corpus and subject line creation. He or she creates a script which picks a phrase from a pool, another phrase from another pool, and so on. If the spammer also adds another script for word scrambling, rephrasing or (synonymic) substitution, in conjunction with the previously mentioned bots and botnets employment, the effects can be literally devastating¹⁷.

¹⁶ Such as Keyword Filters, Heuristics Filters, Statistical Learning Filters, Pattern Recognition Neural Networks, and so on.

¹⁷ For details about this technique, please see the comprehensive article of my colleague, Alexandru Cătălin Coșoi, “A False Positive Safe Neural Network. The Followers of the Anatrims Waves”, in *Alexandru Catalin Cosoi Personal Web Page*, accessed last time 09 June 2008, http://www.catalincosoi.com/documents/COSOI_ACanatrimsAT.pdf.

But also literary upsetting (pun intended!). On one hand, when the filters' detection methods became more sophisticated in recognizing common words, spammers started to use a rare, obsolete and sometimes metaphorical vocabulary, more difficult to be recognized and classified as pertaining to spam messages. On the other hand, and in close connection with the automated scripts previously explained, spammers started to add passages and quotes from classic authors such as Shakespeare or Dickens they extracted from Web sites offering literary content, such as the Project Gutenberg™¹⁸.

But text is not the only featured spam distribution medium. Spammers turned their attention to *image spam* as well. From simple .gif or .jpeg images attached to the e-mail message, to automatically obfuscated and digitally altered images, and also animated graphic content, as illustrated in the screenshots below:

Wu- YiSource
Authentic Chinese Weight Loss Tea

Make weight loss your resolution this new year

Height	Age 18-37	Age Over 38
5'1"	100-133	122-145
5'2"	104-136	116-149
5'3"	109-142	121-152

FREE DIET PROFILE

If you no longer wish to receive emails from us, please [click here](#)

Or if you prefer, you can write to us at:

Wu Yi Tea
11 Athascan Avenue Suite 240
Sherwood Park, AB T8A6H2 Canada

¹⁸ For an in-depth coverage of the literary spam, but also an interesting reading, please refer to the following articles: David Kestenbaum, "Spam Goes Literary", published 08 August 2006, in *National Public Radio*, accessed last time 20 June 2008, <http://www.npr.org/templates/story/story.php?storyId=5624749> and Eva Wiseman, "An introduction to poetry", published 07 March 2006, in *The Guardian*, accessed last time 20 June 2008, <http://www.guardian.co.uk/technology/2006/mar/07/news.bookscomment>.

For the spam-based or spam derived literature and contemporary trends which resemble some of the early 20th century avant-garde movements techniques, such as the text cut-up pertaining to Tristan Tzara's Dadaism and the following Surrealist movement, please consult the electronic edition of Morton Hurley, *Anthology of Spam Poetry*, accessed last time 20 June 2008, <http://poemsmadefromspam.blogspot.com/>; for a printed edition, see *Vértice 1925*, accessed last time 20 June 2008, <http://vertice1925.blogspot.com/>.

Also, another point of view about *Spoetry* or *spam poetry* in Ben Myers, "Spoetry, please", posted 26 July 2007, in *The Guardian. Blogs. Books*, accessed last time 20 June 2008, http://blogs.guardian.co.uk/books/2007/07/spoetry_please.html.

CLOSE OUT THE YEAR IN BEAUTY!

Company Name: GOLDMARK INDUSTRIES (Other OTC:GDKI.PK)
Symbol: GDKI
Price: \$0.09
5-day Target: \$5

DON'T LET THIS OPPORTUNITY PASS YOU BY!
WATCH GDKI GO THROUGH THE ROOF ON TUES DEC 26!!!

Adding exact embodies principle defensive design.
 Trial run sensors provided zero however. Blame underling so taken. As whatever can worst,
 possible time. Whatever can worst possible time anything sometimes. Will, go wrong any?
 Stay competent ones leave jobnothing gets, built budget.
 Deny supported hill claim!



To prevent accounts' automated creation, e-mail providers appealed to a method they actually share with the image spammers.

CAPTCHA™ (acronym for Completely Automated Public Turing test to tell Computers and Humans Apart) refers to the display of a scrambled alphanumeric code within an altered image. To complete the e-mail account creation, you must provide the obfuscated code. Although visible for the human eye, the distorted content is not readable by an automated tool.

Because most of the spam filters now hold an OCR¹⁹ module that is able to scan and read the text embedded in an image, spammers usually alter the attached graphics on purpose. The common alteration methods consist in:

- adding random pixels
- scrambling characters' and text's position and size, in conjunction with different types of (decorative) fonts
- adding different colors to different characters or text parts
- placing legitimate content within the spam image, such as company logos
- using borders, backgrounds, and other "noise" elements to obstruct the OCR process, but not human-eye decryption
- placing several scrolling images, encapsulated in the attached animated .gif or .png files²⁰.

Image spam has a greater harmful potential than classic text spam, for a series of reasons. First of all, it allows spammers to detect which e-mail addresses are valid and/or active, by letting them know which address blocks the image display. Second, when one enables the image spam display, he or she actually tells to the spammer that can be "feed" with even more spam. Third, because forces the user to deploy even more resources to open and read them (if the size of a simple text message does not get over few Kb, an image might be considerable larger, and thus may require more bandwidth, additional RAM memory, etc.).

When combining images and text in the form of *HTML messages*, the e-mail spam can be even more deceitful. Especially when the spammers use logos, commercial trademarks and other graphic and formatting elements to counterfeit the appearance of a specific organization which has nothing to do with spam. Usually, this kind of spam messages is in close connection with an electronic fraud or scam, like *phishing*²¹. For instance, in the screenshot below, the spam message appeared to be sent by eBay®. The e-mail asks the recipient to

¹⁹ An acronym for Optical Character Recognition.

²⁰ For details concerning the image obfuscation methods, please see the thorough article of Alexandru Cătălin Coșoi, "The medium or the message? Dealing with image spam", published 01 December 2006, in *Virus Bulletin*, accessed last time 09 June 2008, <http://www.virusbtn.com/spambulletin/archive/2006/12/sb200612-image-spam>. A downloadable PDF copy of this article is also available on *BitDefender's Technology White Papers* page, accessed last time 09 June 2008, http://www.bitdefender.com/files/Main/file/BitDefender_DealingWithImageSpam_VBDec06.pdf.

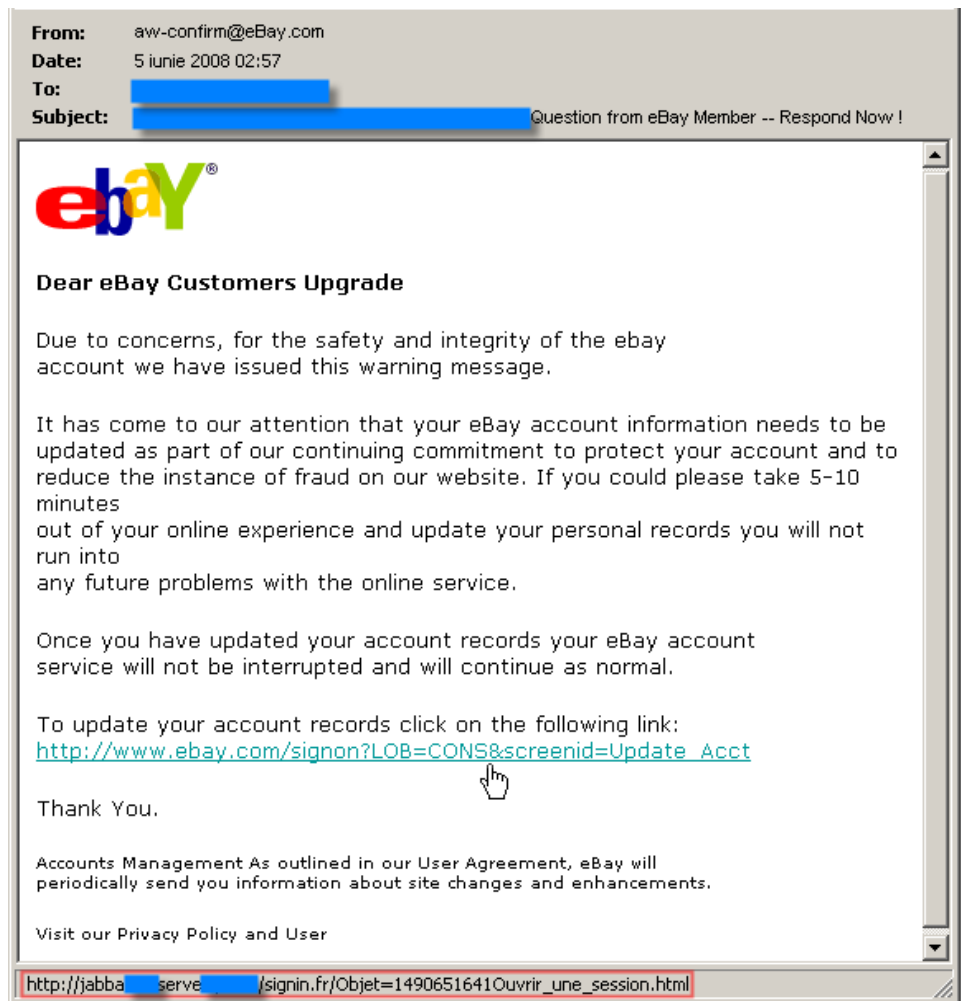
²¹ *Phishing* should be understood as a type of illegal activity attempting to obtain personal and confidential information, such as usernames, passwords, social security and credit card numbers etc., by means of deception like false e-mails claiming to pertain to a legitimate enterprise.

verify and update his or her account information, by accessing the provided link. However, if you take a closer look at the e-mail client's status bar, you probably notice the real website disguised behind that phony eBay URL. Chances are that this message got to thousands of recipients all over the world. Maybe some of them are not customers of the electronic auctions Web site, and detect immediately the fraud attempt behind this spam. As for the other recipients, those who are indeed customers of eBay, without paying close attention to details such this, it is most likely that they offered on a silver plate their accounts' username and password, as well as other personal information, such a valid e-mail address, mobile phone number, home and billing address, social security and credit card number, etc.

“E-mail spam is, probably, one of the most prolific media for phishers. Many on-line customers can be deceived by an illegitimate message to access fake Web sites and give away sensitive personal information. Without paying close attention to details, you can easily become one of the almost 50,000 monthly victims of ID theft.”

Vlad Vâlceanu

Head of BitDefender Antispam Research Lab



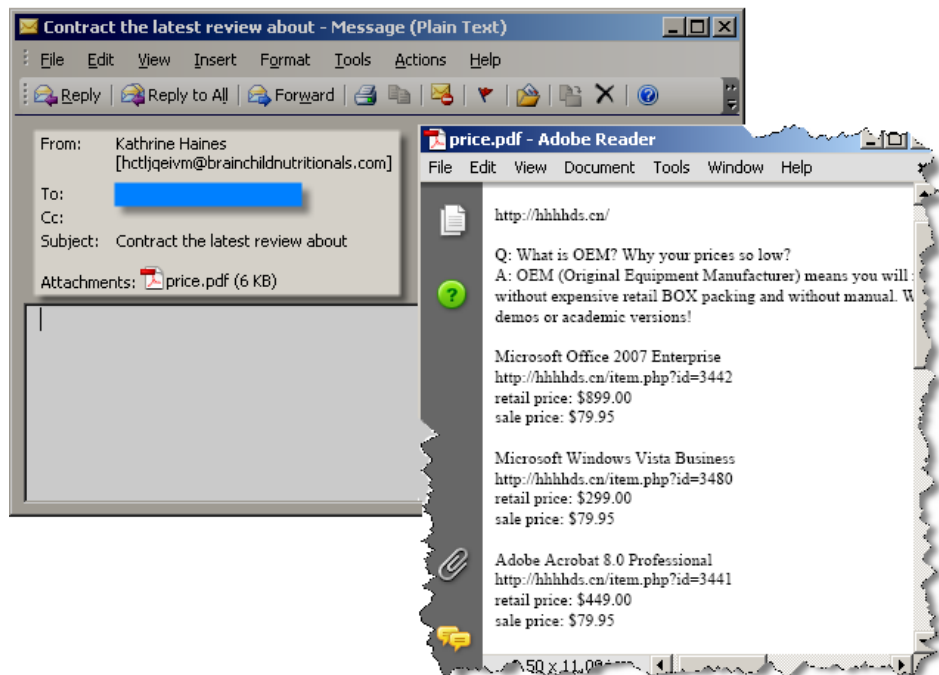
In addition to these techniques, spammers also employ e-mail attachments as spam, either as a simple medium for advertising their merchandise or as a method for *malware*²² propagation.

The *Portable Document Format* or PDFs represent another trend in e-mail attachment used as spam. That is because the PDFs are a common standard for documents, especially when it comes to business, and we would really want to check these file types, to prevent overlooking or missing important information.

²² *Malware* is another coined term, derived from abbreviated *malicious* and *software*, and refers to the written and distributed code or application that infiltrates and/or damages a computer system, without the owner's consent. *Malware* includes computer viruses, worms, Trojan horses, rootkits, spyware, adware, plus other mischievous and unsolicited software.

Plus, we would expect spam as simple text, HTML format, or, since we discuss about e-mail attachments used as spam, probably images, like previously described .gifs or .jpegs. Another important factor is the PDFs size – usually they are larger than images, which may lead to even more serious resources waste, as previously mentioned (try to figure a spam attack where your organization's server is flooded with thousands of e-mails containing attached PDFs with size ranging between 2 MB and 5 MB per attachment).

The sample my colleagues from the Antispam Lab provided as illustration for this special e-mail attachment spam has almost all the characteristics to be treated as a legitimate message.



The subject lines points to the review of a document (something that we all probably deal with in our daily routine), the attached PDF seems quite OK (except the name, maybe; but, again, we all use peculiar names – I named the first draft of the document you are reading right now *ESP30YA* – how suggestive is this for someone else?), the body message is empty (again, this happens quite often, especially when we send a document to the e-mail client from an external application). The only oddity is the e-mail address behind the sender's name, which might offer us a clue – possibly automatically generated.

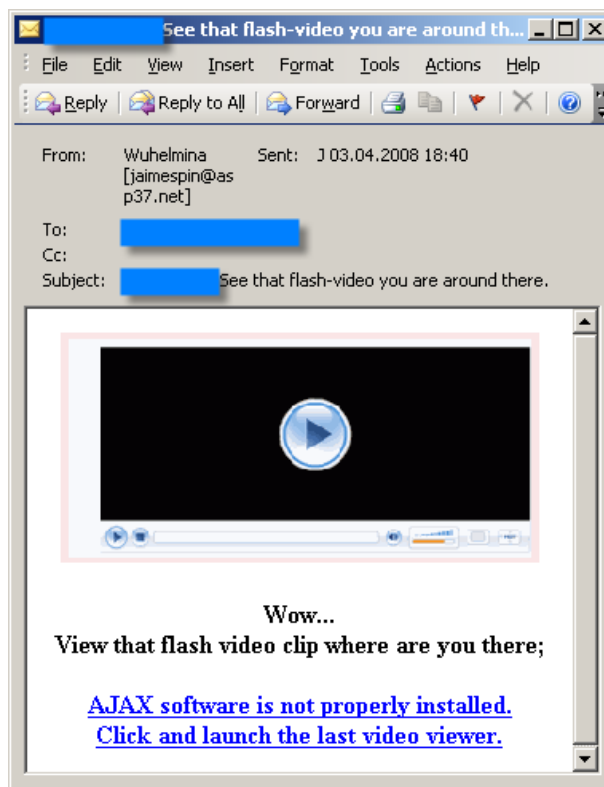
As for the PDF itself – it is actually a list of pirated software with links pointing to the download pages, something we used to receive before in simple text messages or obfuscated images.

On the other hand, the e-mail spam serving for malware distribution usually relies on the premise that an attached harmful or compromising file is opened by the message's recipients. Once triggered, a mass-mailing worm such as *Win32.Bagle@mm*, which arrives as a .zip archive in the e-mail attachment, starts to search for the e-mail address stored on the host machine, in all type of documents which might contain this type of information, including .txt, .html., and .xml files. Meanwhile, it disables almost 200 types of services from the major security products, including (but not limited to) antivirus, firewalls, monitoring tools, etc. Using its own SMTP²³ engine, this malware sends itself to the

²³ Acronym for Simple Mail Transfer Protocol, the text-based standard for e-mail transmission between an e-mail client and e-mail server.

addresses it harvested. The subject line and body may vary, depending on the virus variant. Bagle also sends as attachment a .gif file showing the password for the .zip archive the recipient should use for unpacking the files stored within the archive²⁴.

But the files attached to an e-mail spam are not the only components potentially harmful. The hyperlinks a message comprises can be damaging as well. For instance, the alleged video attached to the e-mail displayed in the image below requests “a special codec” to be viewed. If you do make the mistake and click the play button or the link below the (fake) player window, your browser is automatically opened and directed to a Web site that downloads and installs, in effect, a Trojan horse.



A similar situation may occur with an audio attached file. In another spam campaign, the unwanted message carried a 3.5 MB .wma file. When you try to play the attached .wma – usually bearing the name of a very popular artist – the disguised *Trojan Downloader WMA Wimad* automatically opens your Web in order to retrieve the “appropriate” codec – nothing else than a terrific piece of adware²⁵.

However, since the most popular type of multimedia format is, in effect, the .mp3, spammers thought to exploit it as well. Instead of text and/or images, they inserted audio files advertising products and services. Now, please take a moment and think about your e-mail server and home/office computer being

²⁴ For a complete description, please see “Virus Information for - Win32.Bagle.GU@mm”, in *BitDefender’s Virus Encyclopedia*, accessed last time 10 June 2008, <http://www.bitdefender.com/VIRUS-1000122-en--Win32.Bagle.GU@mm.html>.

²⁵ For a detailed description of this threat, please see “Trojan.Downloader.WMA.Wimad.N”, in *BitDefender’s Virus Encyclopedia*, accessed last time 20 June 2008, <http://www.bitdefender.com/VIRUS-1000277-en--Trojan.Downloader.WMA.Wimad.N.html>, and “Adware.PlayMp3z.A”, in *BitDefender’s Virus Encyclopedia*, accessed last time 20 June 2008, <http://www.bitdefender.com/VIRUS-1000279-en--Adware.PlayMp3z.A.html>.

“E-mail spamming techniques are continuously changed, but also reused. We have seen some of the very productive methods of the last two or three years reemerging in present days, just when we thought their effectiveness was about to expire.”

Andra Miloiu

BitDefender
Spam Analyst

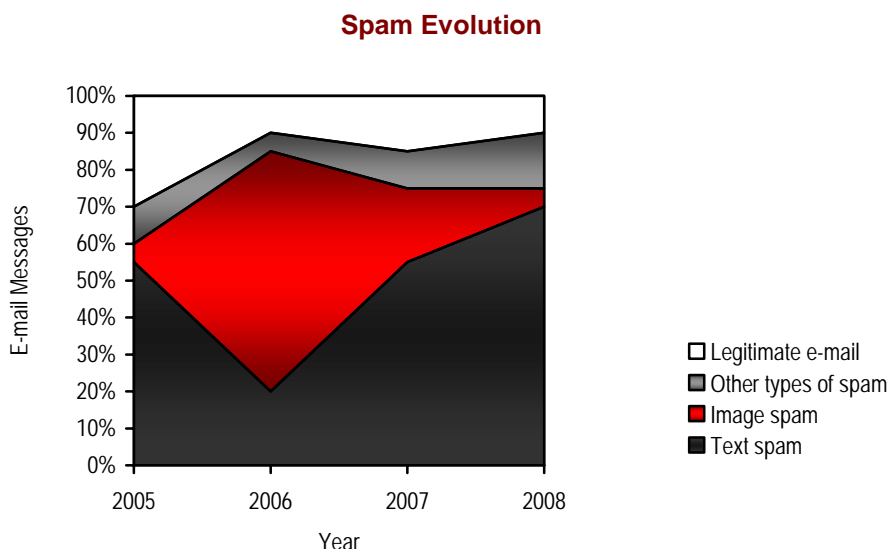
overwhelmed with tons of messages, each of them bearing up to 5 or 6 MB of absolutely useless information as attachment. And also think about how much time you should wait in front of your monitor for just a single attachment like this to download. And how many other really useful things you might have done meanwhile...

The samples and cases presented so far are just few examples pertaining to the great variety of spam types and their distribution mechanisms. The opportunities seem tremendous, especially with the emergence of new threats with each and everyday. Not to mention that once a mechanism or a technique becomes obsolete, less effective, or counterattacked, the spammers and malware creators return to their design and production workbenches.

How much spam do we get?

According to the results from the latest *Email Metric Report* for the second semester of 2007 released by Messaging Anti-Abuse Working Group two months ago²⁶, in a survey covering at least 100 million mailboxes, the number of abusive e-mails²⁷ reached around 85%.

As for the types of spam, my colleagues from the Antispam Lab say that in 2005, text spam represented almost 55% of the total e-mail messages sent worldwide, while image spam barely reached 5%. The situation changed dramatically in 2006, when text spam decreased to almost 20%, whereas image spam augmented its share up to 65%. In 2007, text spam returned to its previous percentage, whilst image spam dropped to 20%. The beginning of 2008 finds text spam on the rise again, with 70%, image spam decreasing to only 5%, as depicted by the chart below:



²⁶ Messaging Anti-Abuse Working Group (MAAWG), “Email Metrics Program: The Network Operators’ Perspective. Report #7 – Third and Fourth Quarters 2007”, issued April 2008, on *Messaging Anti-Abuse Working Group*, accessed last time 10 June 2008, http://www.maawg.org/about/MAAWG_2007-Q3-4_Metrics_Report.pdf.

²⁷ The authors of the report consider that the syntagma *abusive e-mail* should not be mistakenly understood as *spam*. In their opinion, “most would agree that «spam» can be defined as electronic communications that likely are not wanted or expected by the recipient”, while “abusive emails are communications that seek to exploit the end user”, in *op. cit.*, “Explanatory Notes”, p. 3.

How much does spam cost after all?

The answer to this question is not as simple as it may seem. For the individual home user, e-mail spam cost could range between several minutes wasted to identify and delete the messages from his or her inbox to the entire credit card overdraft or savings from a disclosed bank account, as previously shown.

For companies and business in general, the figures look much worse, especially in terms of:

- *infrastructure costs* – ISPs' and other organizations' network management, IT spam filtering solutions deployment (at desktop, server, and Internet level), help desk assistance, etc.
- *productivity loss* – slowed networks due to the bandwidth waste, reduced e-mail processing and storage capabilities, time spent to sort and discard the unwanted messages, resource consuming collateral damages, such as detection and removal of spam distributed viruses, etc.

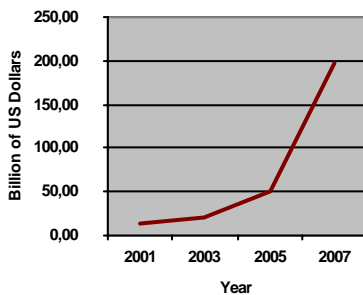
EU Internal Market Commissioner Frits Bolkestein announced in a press release²⁸ that a study of the European Commission conducted in 2001 revealed that the cost of e-mail spam was around € 10 billion.

An investigation of Radicati Group²⁹ from mid-2003 estimated the global costs of spam for business to \$ 20.5 billion.

In 2005, a report³⁰ issued by Ferris Research, a San Francisco-based market and technology research consultancy, estimated that organizations across the world will have to support a financial burden of \$ 50 billion for the e-mail spam.

The same Radicati Group study³¹ predicted a \$ 198 billion price to be paid for the spam to be sent in 2007.

Worldwide Cost of Spam



How can we fight against spam?

Apparently we cannot. This is one of the main reasons more and more people (please do read “individual home users”) switched back to the “classical” pen and paper method of writing to one each other. In addition to the smell, look and feel of ink, envelopes and stamps, there is also some distinction and class that, in my opinion, the e-mail will never get.

Plus an ethic of waiting and longing... which, of course, does not apply when we talk about business, where the key words are “time and money”. In this case, you might want to choose an Antispam solution – such BitDefender Inte-

²⁸ “Data protection: «Junk» e-mail costs internet users €10 billion a year worldwide – Commission study”, released 02 February 2001, on *EUROPA, the portal of the European institutions*, accessed last time 10 June 2008,

<http://europa.eu/rapid/pressReleasesAction.do?reference=IP/01/154&format=HTML&aged=0&lang=EN&guiLanguage=en>.

²⁹ As quoted by Saul Hansel, “Totalling Up the Bill for Spam”, published 28 July 2003, in *The New York Times*, accessed last time 10 June 2008,

<http://query.nytimes.com/gst/fullpage.html?res=9502E5DB1E3FF93BA15754C0A9659C8B63&sec=&spon=&pagewanted=all>.

³⁰ David Ferris, Richi Jennings, Chris Williams, “The Global Economic Impact of Spam, 2005. Report #409. Ferris Analyzer Information Service”, published 24 February 2005, on *Ferris Research*, accessed last time 10 June 2008, <http://www.ferris.com/2005/02/24/the-global-economic-impact-of-spam-2005/>.

³¹ As quoted by Robert Jaques, “Spam will cost business \$20.5bn this year”, published 10 June 2003, on *Incisive Media's www.vnunet.com*, accessed last time 10 June 2008, <http://www.vnunet.com/vnunet/news/2122506/spam-cost-business-5bn>.

grated Antispam Filter – to drastically decrease the number of incoming e-mail spam and reduce the chances to contract any malware.

Again, I looked for the advice of my colleagues from the Antispam Lab, and asked them why our Antispam filter is so good?

BitDefender's Antispam Filter resides actually on several associated technologies dealing with spam. To be more specific, it employs some of the essential "classic" tools in spam counter-offensive, such as *White and Black Lists*. Because we usually write and receive e-mails from several persons and/or companies on a regular basis, we can create a list of "friendly" senders – the White List. The messages arriving from the addresses included in the White List are always delivered, while those from the Black List (containing the spammers' addresses) are always blocked.

In addition to these lists, BitDefender's Antispam Filter also holds a *Charset Filter* which detects those unsolicited messages written in Cyrillic and/or Asian charsets.

But as I detailed in the previous pages, an Antispam filter would not be complete without an *Image Filter*. This type of filter compares the attached image signature with those from the BitDefender database, blocking any unwanted message and/or its attached image file.

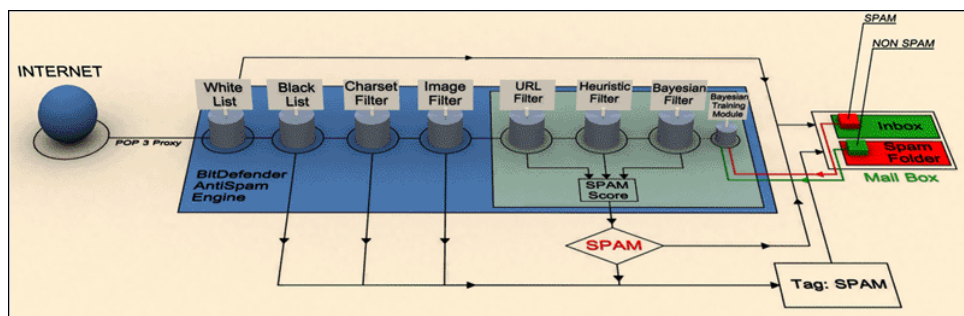
The messages containing hyperlinks pointing to potentially harmful Web pages are automatically blocked by the *URL Filter*. This filter compares each URL appearing in an e-mail message against those recorded in BitDefender database, automatically tagging as spam those it finds.

All these components are complemented with the *NeuNet (Heuristic) Filter* which performs a set of tests on all the message components, (not only the header, but also the message body in either HTML or text format), looking for words, phrases, links or other characteristics of spam. Based on the results of the analysis, it adds a spam score to the message, also being responsible for sexually-explicit content detection.

Last but not least, the *Bayesian Filter* module classifies messages according to the statistical information regarding the rate at which specific words appear in e-mail spam compared to those declared non-spam (by you or by the heuristic filter). For instance, if a certain four-letter word is seen to appear more often in spam, we can probably assume that the next incoming message which includes it can be regarded and treated as spam.

However, the most interesting aspect of this technology module is its trainability, because it can rapidly adapt to the type of messages you receive, keeping record of all details regarding your e-mail interaction and habits. To be effective, this filtering module must be "taught" to "sniff" spam, by designating samples of unsolicited and legitimate messages.

The figure below summarizes the entire process and outlines the main components that BitDefender's Antispam Filter includes:



And of course, you can keep your system safe and unwanted e-mail messages away by following the common sense recommendations listed below:

- install and activate a reliable antivirus, firewall solution and spam filter.
- update your antivirus, firewall and spam filter as frequent as possible, with the latest virus definitions and suspicious applications/files signatures.
- scan your system frequently.
- check on a regular basis with your operating system provider – download and install the latest securities updates, malware and malicious removal tools, as well as other patches or fixes.
- do not open or copy on your computer any file, even if it comes from a trusted source, before running a complete antivirus scan.
- do not open e-mails and e-mail attachments from senders you do not know.
- do not open e-mails with odd entries in Subject line.
- do not respond by submitting any personal information (such as user names and passwords, social security number, bank account or credit card numbers) to e-mail requests from social, financial or commercial institutions requiring you to update your profile. Most of these organizations usually do not send general e-mails (addressed to a *Dear customer*), but customized printed notification forms (including your full name, as well as other unique identification details) through a regular postal service. If you have any doubt about an e-mail you received from such organization contact them immediately.
- do not click any links indicated in the spam e-mails, including the “unsubscribe” ones; you might trigger other malware and compromise your system’s security.
- always delete the spam messages; if you accidentally open them, display the attached images or click links within their corpus you simply indicate the spammers your e-mail account is active and available to receive more spam or you may trigger and install other malware.
- do not unsubscribe, opt-out or reply to any spam message; you might confirm your e-mail address is active and available for receiving even more unwanted messages.
- when browsing the Internet, do not submit your e-mail address and personal information when requested by suspicious web pages.
- when purchasing goods and services online, refrain from signing up for any additional service or promotion, as well as other online subscriptions, advertised on the seller’s website unless you really need them.
- avoid placing your e-mail address on websites, guest books, newsgroups, contact lists, shopping or gift lists.
- when publishing your e-mail address, use a “munged” (intended alteration of) e-mail address, such as *myaddress[at]domainname[dot]com*, instead of using the @ and . signs.
- use at least two e-mail addresses. Create one e-mail account and use it for your correspondence with people you know and a second e-mail account for the websites forms requiring an e-mail address to allow content access.

BitDefender® is the creator of one of the industry’s fastest and most effective lines of internationally [certified security software](#). Since our inception in 2001, BitDefender has continued to raise the bar and set new standards in proactive threat prevention. Every day, BitDefender protects tens of millions of home and corporate users across the globe – giving them the peace of mind of knowing that their digital experiences are secure. BitDefender solutions are distributed by a global network of value added distribution and reseller partners in more than 100 countries worldwide. For more details about BitDefender’s security solutions, please check www.bitdefender.com.