*By Alina P*

Published: 2007-06-11 18:34

# Introduction to Antispam Practices

According to a research conducted by Microsoft and published by the Radicati Group, the percentage held by spam in the total number of emails sent daily has been constantly growing since 2005. As a result, spam is expected to represent 77% of emails sent worldwide by 2009, amounting to almost 250 billion unsolicited emails delivered every day.

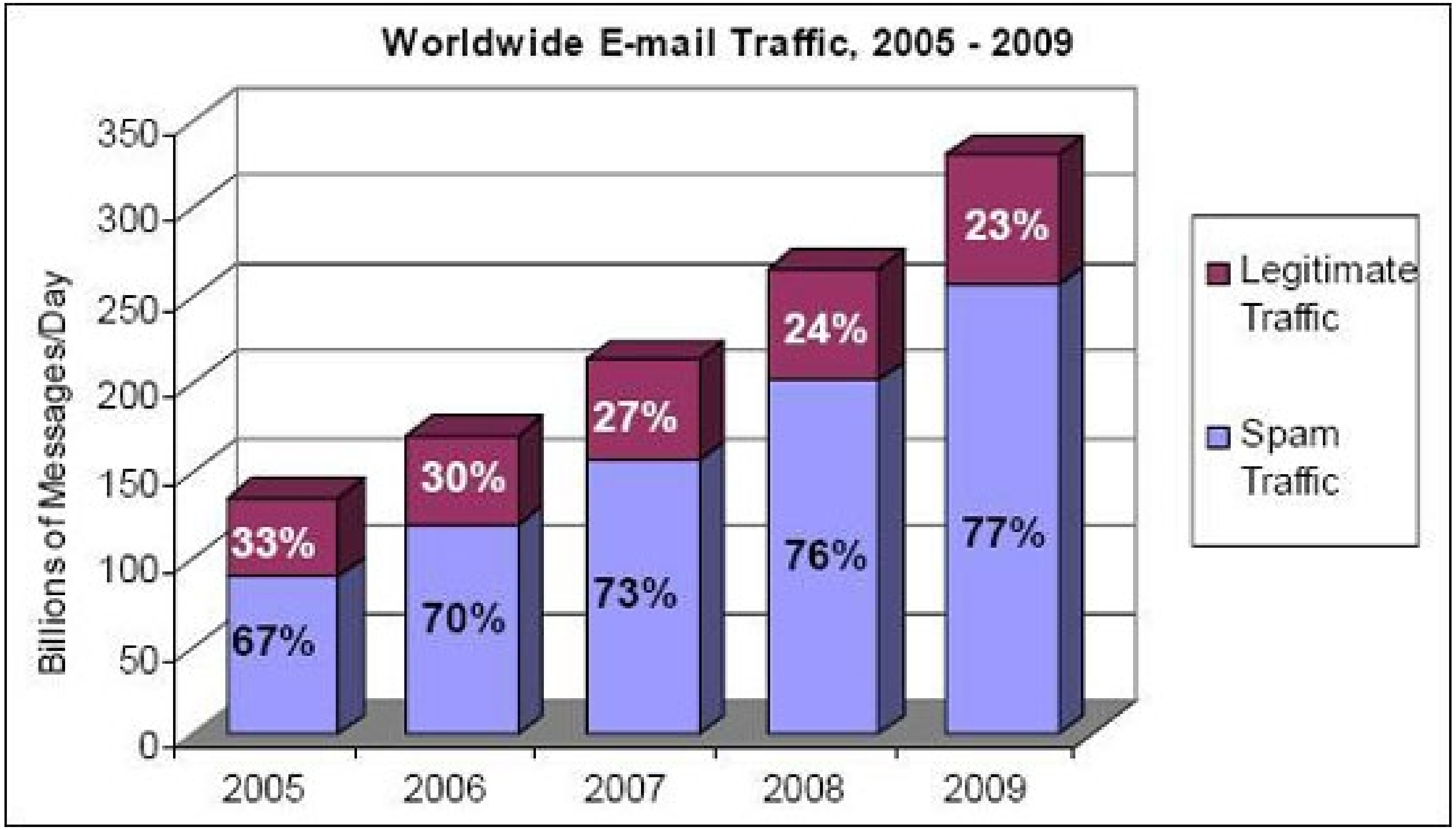


Figure 1. Legitimate Email Traffic vs. Total Spam Traffic, 2005-2009

In a world where spam is bound to hold such an important position, methods of preventing it should also be given an increasing importance. Some of the easiest and most widely used prevention methods are host control solutions, Antispam applications and user education.

Host control is an easy way to ensure only valid emails reach end-users(TM) inboxes. Some well known methods are SPF (Sender Policy Framework), IP/email address-based lists (blacklisting, whitelisting and graylisting) and DKIM (Domain Keys Identified Mail Signature).

# SPF

The Sender Policy Framework (SPF) is an open standard specifying a technical method to prevent sender address forgery. It protects the envelope sender address, which is used for message delivery. The envelope sender address is used during the transport of the message from mail server to mail server, usually not displayed to the user by mail programs.

Using this method, domains can publish details of their mail sending policy (called SPF records) on Domain Name System (DNS) servers. By using SPF checks to validate sender addresses, you can successfully prevent spam and back-scatter emails. Although an effective method of authentication and spam prevention, not all MTAs and ISP providers support SPF checks at this time.

## DKIM

Domain Keys Identified Mail Signature is an authentication method implemented by Yahoo and supported by Google, Cisco and Sendmail and has considerable chances of becoming the standard authentication method. It offers almost end-to-end integrity from a signing to a verifying Mail transfer agent (MTA). In most cases the signing MTA acts on behalf of the sender, and the verifying MTA on behalf of the receiver.

DKIM implies using a key pair consisting of a public key and a private one as follows: the signing MTA generates a public key, which is published in DNS, and a private key, used to digitally sign all the sent email messages. The verifying MTA retrieves the public key and compares it to the digital signature of the received email. If the key pair is a match, then the email is legitimate and is delivered to the receiver(TM)s mailbox.

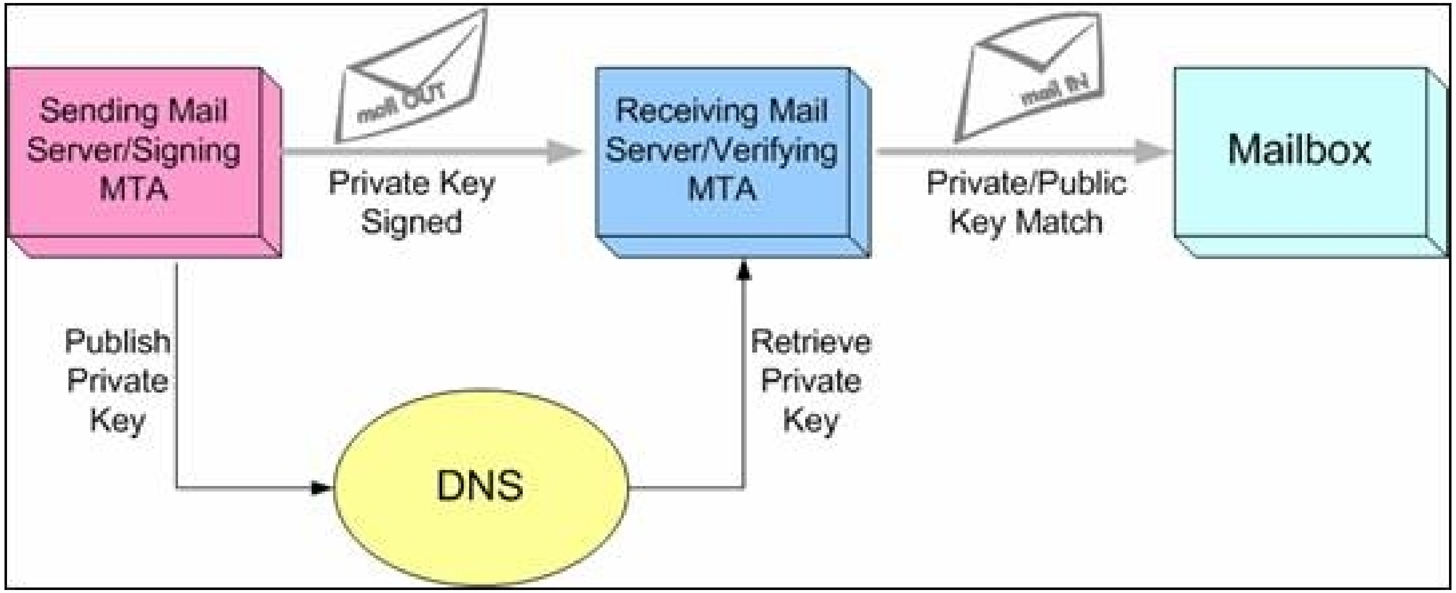


Figure 2. DKIM Compliant Email Flow

The wide use of DKIM can force spammers to show a correct source address. Thus other filtering techniques (such as collaborative databases) can be used to detect spam more reliably. Therefore, DomainKeys can make it easier to identify emails known to be legitimate and need not be filtered. The main benefit in such a case would be saving time and system resources.

The main disadvantage of DKIM is that email messages can be significantly modified in certain situations (e.g. when being forwarded by list severs), causing the signature to be invalidated and the message to be rejected. A solution to this issue would be combining DomainKeys with SPF, because SPF is immune to modifications of the email data.

## Blacklisting

DNS blacklisting is the practice of comparing the routing addresses of incoming e-mails to a list of servers that spammers are suspected to use. Blacklists are either public (free of charge) or private and usually contain IP addresses of open-relay servers, open proxies and ISPs with no spam filtering. If an e-mail appears to be from a blacklisted server, it is blocked, usually with an error message for the recipient.

 

The main advantage of using blacklist is their convenience. Instead of installing and/or training a product to block spam, one can use blacklists to perform the needed email filtering.

On the down side, blacklists never provide an easy way of tracking what and how much is actually being blocked. Moreover, they are known to be highly inaccurate. A given block of addresses may appear on a blacklist, simply because one computer in that block may be a spammer. This causes everyone in the block to be unable to send or receive emails.

## Whitelisting

Whitelists are the exact opposite of blacklists. Instead of filtering possible spammers, they keep track of secure IP addresses or email addresses. They are lists of contacts that the user deems are acceptable to receive email from. An email whose sender is on such a list will be accepted with no further filtering.

Internet service providers are known to use whitelists to filter emails to be delivered to their customers. ISPs receive requests from legitimate companies to add them to the ISP whitelists. Companies either pay for a time period to be allowed to email their customers or the companies pay per complaint received by the ISP from their customers.

The main disadvantage of such lists is that, when set incorrectly, they may reject legitimate emails. However, such incidents tend to appear at end-user level, not at server or ISP level.

## Graylisting

Graylisting is basically a request to have the email resent, after being temporarily rejected by a Mail Transfer Agent (MTA). The sender IP and the recipient of all unknown senders are saved by the MTA which then returns a temporary error. Based on the fact that spammers and spamming scripts do not resend messages, only valid servers will react to this temporary error.

Greylisting requires no additional configuration made by end-users. If the server utilizing greylisting is configured accordingly, the only effect on end users will be a delay on the first message from a given sender. Moreover, far less system resources are used on sending temporary errors than on using Antispam software.

Many MTAs cannot differentiate at this time between a temporary and a permanent error, thus important emails may be lost. This can be prevented however by using whitelists. Greylisting delays much of the mail from non-whitelisted mail servers - not just spam - until typical patterns of communication are recorded by the greylisting system. Therefore, it cannot always prove to be an optimal solution.

## Antispam applications

Software products specialized in preventing spam can be used by end users and by MTA and ISP administrators as well. Open source or commercial, there is a wide range of such solutions to choose from, depending on available resources.

Besides the available host control methods, the filtering system provided by Antispam solutions also embeds content filtering. Thus, they can implement lists of keywords or images most likely to be used by spammers.

An important feature of certain content filtering chains is their use of Bayesian filters. Such filters divide email into spam or legitimate based on the probabilities of certain words or phrases to occur in both types of messages. For example, most email users will frequently encounter the word œViagra• in spam email, but will seldom see it in a legitimate email. However, the filters do not know these probabilities in advance. To train the filter, users must manually indicate whether a new email is spam or not. Based on all words in each training email, the filter will assign probabilities to each word in its database, both for spam and legitimate messages.

The main advantage of Bayesian spam filtering is that it can be trained on a per-user basis. The training, however, is time consuming. Moreover, incorrectly identifying emails as spam or legitimate may lead to having to restart the whole training process.

## Educating End-Users

Although sometimes overlooked, this aspect is one of the most important aspects of fighting spam. Incorrectly handling existing spam messages often leads to receiving more unsolicited emails.

The most common example is represented by unsubscribe links included in spam messages. A large number of those receiving spam are tempted to click on such links contained by unsolicited emails. Such an action will let the spammer know there is a live person with a valid email address. The email address will be tagged as a good one and can later be sold to other spammers. In a worse case, the link could lead to a website that installs a virus or spyware, leading to severe security issues.

According to a study conducted by Radicati Group and Mirapoint in March and April, 2005, 55% of email users follow the unsubscribe links in spam messages. The correct procedure in such cases is to permanently delete such messages.

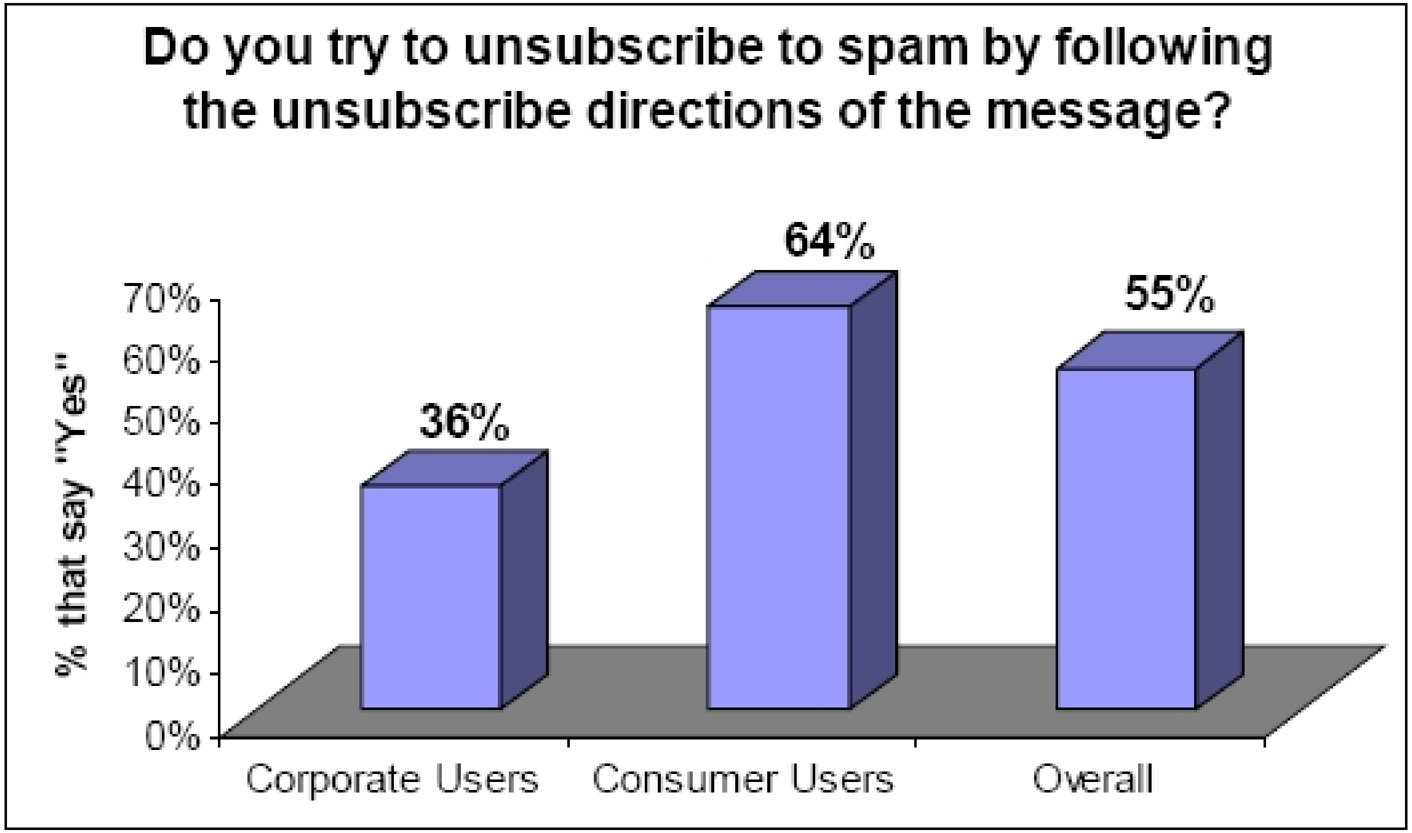


Figure 3. Users that Follow the Message(TM)s Directions to Unsubscribe

Making sure all users know how to treat spam is therefore the first step to be taken in order to reduce this phenomenon. It should also be doubled by research and development in technology and better laws for this field being made available worldwide. However, without proper user knowledge, technology is almost useless and laws cannot be applied; therefore the troubling predictions on future spam expansion may easily come true. Permanent efforts to educate users are not a full proof defense against unsolicited mail, but represent the only optimal available protection.

## Resources:

**[http://www.axigen.com](http://www.axigen.com)**

**http://www.radicati.com/reports/whitepapers.asp#**
**http://antispam.yahoo.com/domainkeys**
**http://www.openspf.org/**