eScan Anti-Spam

White Paper

Document Version (esnas 14.0.0.1)
Creation Date: 19th Feb, 2013

Preface

The purpose of this document is to discuss issues and problems associated with spam email, describe available technologies to fight spam and explain how eScan unique technology can fight and prevent spam.

The Email Spam Problem

What is spam? An Unsolicited commercial email, commonly referred to as spam, is growing rapidly and is finding its way to desktops both in the home and in the office. The major impact of spam is on ISPs that must cope with increasing amount of email traffic. Businesses are also suffering from spam attacks that affect their infrastructure and productivity. **Spam** is the use of electronic messaging systems to send unsolicited bulk messages, especially advertising, indiscriminately. While the most widely recognized form of spam is email spam.

Spamming remains economically viable because advertisers have no operating costs beyond the management of their mailing lists, and it is difficult to hold senders accountable for their mass mailings. Because the barrier to entry is so low, spammers are numerous, and the volume of unsolicited mail has become very high. In the year 2012, the estimated figure for spam messages is around seven trillion. The costs, such as lost productivity and fraud, are borne by the public and by ISPs, which have been forced to add extra capacity to cope with the deluge. Spamming has been the subject of legislation in many jurisdictions.

A person who creates electronic spam is called a *spammer (source Wikipedia)*

How to stop spam using eScan Anti-spam technology?

There are several techniques to fight spam; none of them can completely prevent it without blocking any legitimate email, in technical term we call it as false positive. However, using combination of techniques spam can be prevented at lowest possible level and preventing any false positives.
Let us look at the methods used to identify and block spam messages

- NILP
- RBL (Real-time black lists)
- IP Reputation
- Reverse DNS Lookup
- Spoofed Sender
- Sender Policy Framework
- Header Check
- Spam Database
- Chinese or Russian Characters
- SURBL
- Content checking
- Image analysis

eScan employs combination of technologies to prevent spam. Most of the methods are available as a part of eScan products.

Here we have explained methods which are part of eScan products:

1. **NILP (Non-Intrusive Learning Pattern)**

This is a revolutionary technology from MicroWorld that works on the principles of Artificial Intelligence to create an adaptive mechanism in Spam and Phishing Control. This technology analyzes each email according to the Behavioral Patterns of the user and then takes an informed decision thereafter. NILP has a self-learning mechanism apart from incorporating regular research feeds from MicroWorld's Server.

How NILP Works?

NILP uses Bayesian Filtering and works on the principles of Artificial Intelligence (AI). It has self-learning capabilities and uses an adaptive mechanism to categorize e mails based on the behavioral pattern of the user. NILP updates itself by using regular research feeds from MicroWorld servers. Whenever a new e mail arrives, NILP analyzes it based on the accumulated learning, and classifies it as ham or spam.

NILP also maintains a database containing the DNA imprints of millions of SPAM e mails, which it keeps updating continuously. It uses the existing DNA imprints in the database to reverse its learning and determine whether a given e mail is ham or spam. In this way, the NILP technology protects the user's inbox from spam and phishing e-mails.

2. **RBL (Real-time black lists)**

This technique, commonly referred to as RBL (real-time black-hole lists), checks the incoming IP address against various Black Lists to verify that the sending server is not listed as an open mail relay that spammers can use to relay their unsolicited emails.

Normally any secure mail server should refuse to relay (send) email from an external sender to anybody outside its domain. This ensures that spammers cannot hijack it and use it to send spam that will look like it is coming from a legitimate mail server.

Unfortunately some system administrators, for whatever reason, fail to configure their mail servers to block such a relay.

eScan comes with a default set of the two most reliable free RBL services. eScan can be configured to use several RBLs at a time for higher accuracy.

3. **IP Reputation**

It is simple. All email must originate from an IP address, and IP reputation can be used to tell if a certain IP Address is responsible for sending Spam or Unwanted Bulk Email (UBE). And it is extremely effective, stopping between 80% of spam emails.

The IP address is a widely known to be a generic spam source. Some networks choose to reject, not deliver, tag as spam or quarantine messages coming from IP addresses that were identified as being under the control of or available for use by spammers. eScan builds the database of received emails to keep a watch of compromised systems. eScan gives grades to IP address based on their reputation, and this reputation are kept in a database. And this database gets updated whenever new entry is added or removed.

4. **Reverse DNS Lookup**

This is not a very good method as it is time consuming. The receiving server performs a "reverse DNS lookup" for the IP address of the incoming connection and check if there is a valid registered hostname associated with it. Reverse DNS lookups take too much time and can result in timeouts and rejection of legitimate emails.

eScan also use HELO lookup, this is a preferred method the receiving server will get the host name of the sending mail server from the SMTP HELO command, perform a simple DNS query and verify that the resolved IP address is indeed the IP address of the incoming connection.

5. **Spoofed Sender**

Email address spoofing is a technique used to send email messages from outside sources masquerading as internal addresses within the organization.

• One example would be to send an email to abc@escanav.com masquerading as xyz@escanav.com.
• Another example would be to send email to abc@escanav.com masquerading as abc@escanav.com. The recipient seems to be receiving a mail from himself.

eScan provides a complete Anti-spoofing solution, blocking email masquerading to be coming from someone within the organization.

6. **Sender Policy Framework**

Sender Policy Framework (SPF) is a validation system designed to prevent spam by detecting spoofing, a common vulnerability, by verifying sender IP addresses. SPF allows administrators to specify which hosts are allowed to send mail from a given domain by creating a specific SPF Record (or TXT record) in the Domain Name System (DNS). Mail Exchanger use the DNS to check that mail from a given domain is being sent by a host sanctioned by that domain's administrators.

7. **Header Check**

Header verification is the process of inspecting the email SMTP header for compliance with standards to make sure that they are not forged by spammers. Also, some spam sending applications insert certain identifiable information info the email SMTP header and other data.

8. Spam Database

The spam database technology extracts phrases from received email and compares against a database of known spam emails. If the spam phrases probability is higher the email is marked as spam. This is a very powerful technology if implemented correctly because it has the potential of blocking spam in real-time.

9. Text Manipulation

Many spam messages use tricks to make it harder for anti-spam tools to textually analyze its content. Text manipulation is a method or replacing certain characters in the spam text with visually similar characters, separating characters so it is difficult to analyze as a whole word, using auditory similar symbols or letter to represent words and parts of words, and more. Some examples are:

- P0RN instead of PORN (zero instead of capital O).
- \/\/arez instead or Warez (\ / \ / used to form the letter W)

eScan can easily detect these kind of spams and prevent them.

10. Chinese or Russian Characters

If Chinese or Russian characters found in an email, eScan has the ability to detect the characters and mark that email as SPAM

11. SURBL

SURBL is a collection of URI DNSBL lists of Uniform Resource Locator (URI) hosts, typically web site domains that appear in unsolicited messages. SURBL can be used to search incoming email message bodies for spam payload links to help evaluate whether the messages are unsolicited. For example, if *http://www.test.com* is blacklisted, then e-mail messages with a message body containing this URI may be classified as unsolicited. eScan extract urls from the emails and check them if they are listed or not. Based on the results it tags email as SPAM.

12. Content checking

This is a traditional method where you specify content e.g. Xanax if this word found in an email it should block that email. This method needs manual feeding of the contents which needs to be blocked or recognized as spam messages.

13. Image analysis

A large portion of spam messages contain visual pornographic content. This content is not just time and resource consuming but can also be offensive. It could even lead to legal liability problems. eScan Advanced Anti-spam service detects spam with links to images. The service's auto-web-crawlers access the web server where the images reside and analyze the images using pattern matching algorithms which detect visual pornographic content.

The solution is smart enough to distinguish nudity from medical or other legitimate images. If the images are identified as pornographic the site will be indexed and added to the URL database.