International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 02 | Feb 2019 www.irjet.net p-ISSN: 2395-0072

Image Spam Detection: Problem and Existing Solution

Anis Ismail[1], Shadi Khawandi[2], Firas Abdallah[3]

[1,2,3]*Faculty of Technology, Lebanese University, Lebanon*

Abstract - *Today very important means of communication is the e-mail that allows people all over the world to communicate, share data, and perform business. Yet there is nothing worse than an inbox full of spam; i.e., information crafted to be delivered to a large number of recipients against their wishes. In this paper, we present a numerous anti-spam methods and solutions that have been proposed and deployed, but they are not effective because most mail servers rely on blacklists and rules engine leaving a big part on the user to identify the spam, while others rely on filters that might carry high false positive rate.*

Key Words: E-mail, Spam, anti-spam, mail server, filter.

1. INTRODUCTION

The internet community has grown and spread widely in a way that not only is it connecting every one of its users into one virtual globe, but also affecting them. Given that the internet is still in an ongoing evolution, states that this virtual community of people (users) is growing and with this growth comes great value, a value of people connected all together in a certain period of time all of the time, now imagine what this could bring forward as a target regarding marketing, advertisement, at the same time it could also hurt such users when such marketing and advertisement are misused, therefore affecting the resource structure of this globe along with its users. Consider a table whose resource structure are its four wooden legs which is able to hold a capacity of 50 kg, now bring a load of 70 kg and you will notice that the table would be crippled and broken, now apply that on the internet community whose resource structure are its communication which is able to hold up to a certain level of bandwidth, if we abuse that level and raise it up the internet community will be crippled and get affected by itself and its users thus costing the whole community a burden which starts from spam.

Spam is flooding the Internet with many copies of the same message, in an attempt to force the message on people who would not choose to receive it, and is also regarded as the electronic equivalent of junk mail. Most spam is commercial advertising and is generally e-mail advertising for some product sent to a mailing list or newsgroup. This is done by the abuse of electronic messaging systems including most broadcast media, digital delivery systems to send unsolicited bulk messages at random. While the most widely recognized form of spam is e-mail spam, the term is applied to similar abuses in other media: instant messaging spam, Usenet newsgroup spam, Web search engine spam, spam in blogs, wiki spam, online classified ads spam, mobile phone messaging spam, Internet forum spam, junk fax transmissions, and file sharing network spam [1]. People who create electronic spam are called spammers [2].

The generally accepted version for source of spam is that it comes from the Monty Python song, "Spam spam spam spam, spam spam spam spam, lovely spam, wonderful spam…" Like the song, spam is an endless repetition of worthless text. Another thought maintains that it comes from the computer group lab at the University of Southern California who gave it the name because it has many of the same characteristics as the lunchmeat Spam that is nobody wants it or ever asks for it. No one ever eats it. It is the first item to be pushed to the side when eating the entree. Sometimes it is actually tasty, like 1% of junk mail that is really useful to some people [2].

E-mail spam is known as unsolicited bulk E-mail (UBE), junk mail, or unsolicited commercial e-mail (UCE), is a subset of spam where in practice it is the sending of unwanted e-mail messages, frequently with commercial content, in large quantities to a random set of recipients. Spam in e-mail started to become a problem when the Internet was opened up to the general public in the mid-1990s. It grew exponentially over the following years, and today is estimated to comprise some 80 to 85% of all the e-mail in the world [1]. Digital image is a representation of a two-dimensional image using ones and zeros (binary). The term "digital image" usually refers to raster images also called bitmap images. Raster images have a finite set of digital values, called picture elements or pixels. The digital image contains a fixed number of rows and columns of pixels. Pixels are the smallest individual element in an image, holding quantized values that represent the brightness of a given color at any specific point.

Typically, the pixels are stored in computer memory as a raster image or raster map, a two-dimensional array of small integers. These values are often transmitted or stored in a compressed form which is the process of encoding information using fewer bits than an unencoded representation would use. Raster images can be created by a variety of input devices and techniques, such as digital cameras, scanners, coordinate-measuring machines, seismographic profiling, airborne radar, and more. Each pixel of a raster image is typically associated to a specific 'position' in some 2D region, and has a value consisting of one or more quantities related to that position. Digital images can be classified according to the number and nature of those samples such as binary, grayscale, color, false-color, multi-spectral, thematic, and picture function [3].

Image spam is a kind of E-mail spam where the message text of the spam is presented as a picture in an image file. Since most modern graphical e-mail client software will render the image file by default by presenting the message image directly to the user, thus it is highly effective at overcoming normal e-mail filtering software where it inputs the e-mail, and as for its output it might pass the e-mail message through unchanged for delivery to the user's mailbox, redirect the message for delivery elsewhere, or even throw the message away.

2. Problem Statement

This paragraph introduces the problem and the effect that spam is having on the internet community and the damage it is producing in addition to the trouble that faces current filters in their quest to efficiently detect the presence of spam within an image and that is due to the new techniques that the spammers are inhibiting in order to come around those filters, also the light will be shed on the need of domain verification mechanism.

2.1 E-mail Impact

The e-mail spam targets individual users with direct mail messages and e-mail spam lists are often created by scanning Usenet postings, stealing Internet mailing lists, or searching the Web for addresses. E-mail spam typically cost users money from their pocket to receive. Many people who have their phone billed for internet usage read or receive their mail while they are being billed for each minute they spend thus, spam costs them additional money. On top of that, it costs money for ISPs and online services to transmit spam, and these costs are transmitted directly to subscribers.

In addition to wasting people's time and money with unwanted e-mail, spam also eats up a lot of network bandwidth. Consequently, there are many organizations, as well as individuals, who have taken it upon themselves to fight spam with a variety of techniques. But because the Internet is public, there is really little that can be done to prevent spam, just as it is impossible to prevent junk mail.

Spamming is still available until now because it does not have an operating cost on the spammer other than the management of the mailing lists, and it is difficult to hold the spammer accountable for his mass mailings. Therefore, given the above feasibility the spammers are numerous, and the volume of unsolicited mail has become very high. The costs incurred such as lost productivity and fraud, have become a liability to the public and the Internet service providers, which have been forced to add extra capacity to cope with the increase of spam.

The e-mail spam has steadily, even exponentially grown since the early 1990s to several billion messages a day. Spam has frustrated, confused, and annoyed e-mail users. Laws against spam have been periodically implemented with some being avoided. Spam averages 94% of all e-mail sent [4].

2.2 Collecting E-mails

An industry of e-mail address harvesting is dedicated to collecting e-mail addresses and selling compiled databases. Some of these address harvesting approaches rely on e-mail addresses that are collected from chat rooms, websites, newsgroups, and viruses which harvest users' address books, while others rely on users not reading the fine print of agreements, resulting in them agreeing to send messages to their contacts. This is a common approach in social networking spam such as that generated by the social networking site Quechup. Much of spam is sent to invalid e-mail addresses. ISPs have attempted to recover the cost of spam through lawsuits against spammers, although they have been mostly unsuccessful in collecting damages despite winning in court [5].

2.3 E-mail Approach

One particularly nasty variant of e-mail spam is sending spam to mailing lists (public or private e-mail discussion forums.) Because many mailing lists limit activity to their subscribers, spammers will use automated tools to subscribe to as many mailing lists as possible, so that they can grab the lists of addresses, or use the mailing list as a direct target for their attacks.

Increasingly, e-mail spam today is sent via "zombie networks", networks of virus or worm-infected personal computers in homes and offices around the globe, many modern worms install a backdoor which allows the spammer access to the computer and use it for malicious purposes. This complicates attempts to control the spread of spam, as in many cases the spam does not even originate from the spammer.

Within a few years, the focus of spamming (and anti-spam efforts) moved chiefly to e-mail, where it remains today. Moreover, the aggressive e-mail spamming by a number of high-profile spammers such as Sanford Wallace of Cyber Promotions in the mid-to-late 1990s contributed to making spam predominantly an e-mail phenomenon in the public mind [6].

2.3.1 E-mail Legalization

From the beginning of the Internet (the ARPANET), sending of junk e-mail has been prohibited, enforced by the Terms of Service/Acceptable Use Policy (ToS/AUP) of internet service providers (ISPs). Even with a thousand users junk e-mail for advertising is not acceptable, and with a million users it is not only impractical, but also expensive. It is estimated that spam cost businesses on the order of $100 billion in 2007. As the scale of the spam problem has grown, ISPs and the public have turned to government for relief from spam, which has failed to materialize.

Pressure to make e-mail spam illegal has been successful in some jurisdictions, but less so in others. Spammers take advantage of this fact, and frequently outsource parts of their

operations to countries where spamming will not get them into legal trouble [5].

2.3.2 Image Spam Work Around

The basic motivation behind image spam is that it is difficult to detect using spam filtering software designed to detect patterns in text in the plain-text E-mail body. Attempts to filter text in image spam are easily defeated because optical character recognition of text in image spam can be prevented using a variety of obfuscation techniques which will not prevent the spam image from being read by human beings.

Obfuscation techniques can include blurring of text outlines, construction of the image from multiple image layers assembled within an HTML e-mail, use of animated image formats, and random noise added to the image (also known as confetti) to prevent the detection of multiple similar images using hash algorithms.

Currently, the surest known countermeasure for image spam is to discard all messages containing images which do not appear to come from an already white listed e-mail address. However, this has the disadvantage that valid messages containing images from new correspondents would be silently discarded [6].

2.4 Kinds of Threats

There are several kinds of methods that are considered harmful to your e-mail and pose a threat to your e-mail and we will list and discuss these threats below

2.4.1 419 Scams

Advance fee fraud spam such as the Nigerian "419" scam may be sent by a single individual from a cyber cafe in a developing country, in which individual receiving such spam could believe in it and would be scammed to give away money or do certain illegal things on behalf of them without him knowing so [7].

2.4.2 Phishing

Spam is also a medium for fraudsters to scam users to enter personal information on fake Web sites using e-mail forged to look like it is from a bank or other organization such as PayPal [7].

2.4.3 Appending

The marketer having one database containing names, addresses, and telephone numbers of prospective customers, can pay to have their database matched against a database containing e-mail addresses, thus purchasing a list of e-mail addresses that match a list of those names to be used later for spam.

2.4.4 Image Spam

Image spam is an obfuscating method in which the text of the message is stored as a GIF or JPEG image and displayed in the e-mail. This prevents text based spam filters from detecting and blocking spam messages. Image spam is currently used largely to advertise stocks [6].

2.4.5 Blank Spam

Blank spam is spam lacking a payload advertisement. Often the message body is missing altogether, as well as the subject line. Still, it fits the definition of spam because of its nature as bulk and unsolicited e-mail.

Blank spam may be originated in different ways, either intentional or unintentionally where blank spam can have been sent in a directory harvest attack, a form of dictionary attack for gathering valid addresses from an e-mail service provider. Since the goal in such an attack is to use the bounces to separate invalid addresses from the valid ones, the spammer may dispense with most elements of the header and the entire message body, and still accomplish his or her goals and blank spam may also occur when a spammer forgets or otherwise fails to add the payload when he or she sets up the spam run. Moreover often blank spam headers appear truncated, suggesting that computer glitches may have contributed to this problem from poorly-written spam software to careless relay servers, or any problems that may truncate header lines from the message body. In addition some spam may appear to be blank when in fact it is not. An example of this is the VBS.Davinia.B e-mail worm which propagates through messages that have no subject line and appears blank, when in fact it uses HTML code to download other files [8].

2.4.6 Backscatter Spam

Backscatter is a side-effect of e-mail spam, where e-mail servers receiving spam and other mail send bounce messages to an innocent party. This occurs because the original message's envelope sender is forged to contain the e-mail address of the victim. A very large proportion of such e-mail is sent with a forged From: header, matching the envelope sender [7].

2.5 Legality

Sending spam violates the Acceptable Use Policy (AUP) of almost all Internet Service Providers. Providers vary in their willingness or ability to enforce their AUP. Some actively enforce their terms and terminate spammers' accounts without warning. Some ISPs lack adequate personnel or technical skills for enforcement, while others may be reluctant to enforce restrictive terms against profitable customers.

As the recipient directly bears the cost of delivery, storage, and processing, one could regard spam as the electronic equivalent of "postage-due" junk mail [9]. Due to the low cost

of sending unsolicited e-mail and the potential profit entailed, some believe that only strict legal enforcement can stop junk e-mail. The Coalition against Unsolicited Commercial E-mail (CAUCE) argues "Today, much of the spam volume is sent by career criminals and malicious hackers who will not stop until they are all rounded up and put in jail."

Spam is legally permissible according to the CAN-SPAM Act of 2003 provided it follows certain criteria. A truthful subject line, no false information in the technical headers or sender address, "conspicuous" display of the postal address of the sender, and other minor requirements, if the spam fails to comply with any of these requirements, then it is illegal. Aggravated or accelerated penalties apply if the spammer harvested the e-mail addresses using methods described earlier.

2.5.1 Introduced Anti-spam Legislations

The Government of Canada has introduced anti-spam legislation called the Electronic Commerce Protection Act at the House of Commons to fight spam. Australia and all the countries of the European Union have passed laws that specifically target spam. In the United States, most states enacted anti-spam laws, which have since been pre-empted by the CAN-SPAM Act of 2003.

2.5.2 Effectiveness

Legislative efforts to curb spam have been ineffective or counter-productive. For example, the CAN-SPAM Act of 2003 requires that each message include a means to "opt out" (i.e., decline future e-mail from the same source). It is widely believed that responding to opt-out requests is unwise, as this merely confirms to the spammer that they have reached an active e-mail account. To the extent this is true the CAN-SPAM Act's opt-out provisions are counter-productive in two ways: first, recipients who are aware of the potential risks of opting out will decline to do so. Second, attempts to opt-out will provide spammers with useful information on their targets. A 2002 study by the Center for Democracy and Technology found that about 16% of web sites tested with opt-out requests continued to spam [9].

2.5.3 Other Laws

Accessing privately owned computer resources without the owner's permission counts as illegal under computer crime statutes in most nations. Deliberate spreading of computer viruses is also illegal. Thus, some common behaviors of spammers are criminal regardless of the legality of spamming. Even before the beginning of laws specifically banning or regulating spamming, spammers were successfully prosecuted under computer fraud and abuse laws for wrongfully using others' computers.

The use of botnets can be perceived as theft. The spammer consumes a zombie owner's bandwidth and resources without any cost. In addition, spam is perceived as theft of services. The receiving SMTP servers consume significant amounts of system resources dealing with this unwanted traffic. As a result, service providers have to spend large amounts of money to make their systems capable of handling these amounts of e-mails. Such costs are inevitably passed on to the service providers' customers.

2.6 Deception and Fraud

Spammers may engage in deliberate fraud to send out their messages. Spammers often use false names, addresses, phone numbers, and other contact information to set up "disposable" accounts at various Internet service providers. They also often use falsified or stolen credit card numbers to pay for these accounts. This allows them to move quickly from one account to the next as the host ISPs discover and shut down each one.

Senders may go to great lengths to conceal the origin of their messages. Large companies may hire another firm to send their messages so that complaints or blocking of e-mail falls on a third party. Others engage in spoofing of e-mail addresses. The e-mail protocol (SMTP) has no authentication by default, so the spammer can pretend to originate a message apparently from any e-mail address.

Senders cannot completely spoof e-mail delivery chains (the 'Received' header), since the receiving mail server records the actual connection from the last mail server's IP address. To counter this, some spammers forge additional delivery headers to make it appear as if the e-mail had previously traversed many legitimate servers.

Spoofing can have serious consequences for legitimate e-mail users. Not only can their e-mail inboxes get congested with "undeliverable" e-mails but in addition to volumes of spam they can mistakenly be identified as a spammer. Also, not only may they receive e-mail from spam victims, but if spam victims report the e-mail address owner to the ISP for example, a naive ISP may terminate their service for spamming.

2.6.1 Theft of Service

Spammers frequently seek out and make use of vulnerable third-party systems such as open mail relays and open proxy servers. SMTP forwards mail from one server to another, mail servers that ISPs run commonly require some form of authentication to ensure that the user is a customer of that ISP. Open relays, however, do not properly check who is using the mail server and pass all mail to the destination address, making it harder to track down spammers. Increasingly, spammers use networks of malware-infected PCs (zombies) to send their spam. Zombie networks are also known as Botnets (such zombifying malware is known as a bot, short for robot). In June 2006, an estimated 80% of e-mail spam was sent by zombie PCs, an increase of 30% from the prior year. Estimated 55 billion e-mail spams were sent each day in June 2006, an increase of 25 billion per day from June 2005 [10].

2.7 Statistics and Estimates

Spam is growing, with no signs of reduction. The amount of spam users see in their mailboxes is just a small evident part, since spammers' lists often contain a large percentage of invalid addresses and many spam filters simply delete or reject "obvious spam".

2.7.1 In Absolute Numbers

The number of spam e-mails have been in rise since the first e-mail spam have been created and Table 1 shows the volume of e-mails per day given certain years

Table 1 - Volume of E-mails per Day [5]

Year	Spam per day
1978	An mail spam advertising a DEC product presentation is sent by Gary Thuerk to 600 addresses, which was all the users of that time's ARPANET, though software limitations meant only slightly more than half of the intended recipients actually received it.
2002	2.4 billion per day
2004	11 billion per day
2005	(June) 30 billion per day
2006	(June) 55 billion per day
2007	(February) 90 billion per day
2007	(June) 100 billion per day

2.7.2 Percentage of the Total Volume of E-mails

More than 97% of all e-mails sent over the net are unwanted, according to a Microsoft security report.

MAAWG estimates that 85% of incoming mail is "abusive e-mail", as of the second half of 2007. The sample size for the MAAWG's study was over 100 million mailboxes.

Spamhaus estimates that 90% of incoming e-mail traffic is spam in North America, Europe or Australasia. By June 2008 96.5% of e-mail received by businesses was spam [11].

2.7.3 Highest Amount of Spam Received

According to Steve Ballmer, Microsoft founder Bill Gates receives four million e-mails per year, most of them spam.

At the same time Jef Poskanzer, owner of the domain name acme.com, was receiving over one million spam e-mails per day [5].

2.7.4 Cost of Spam

A 2004 survey estimated that lost productivity costs Internet users in the United States $21.58 billion annually, while another reported the cost at $17 billion, up from $11 billion in 2003. In 2004, the worldwide productivity cost of spam has been estimated to be $50 billion in 2005 [5]. An estimate of the percentage cost borne by the sender of marketing junk mail is 88%, whereas in 2001 one spam was estimated to cost $0.10 for the receiver and $0.00001 (0.01% of the cost) for the sender.

2.7.5 Source of Spam

Source of spam refers to the geographical location of the computer from which the spam is sent. It is not the country where the spammer resides, nor the country that hosts the spam advertised site. Due to the international nature of spam, the spammer, the hijacked spam-sending computer, the spam advertised server and the user target of the spam are all often located in different countries. As much as 80% of spam received by Internet users in North America and Europe can be traced to fewer than 200 spammers [6].

2.7.6 In Terms of Volume of Spam

A 2009 Cisco Systems report lists the origin of spam by country and is shown in Table 2 as follows [12].

Table 2 - Origin of spam by country

Country	Trillions per year
Brazil	7.7
USA	6.6
India	3.6
South Korea	3.1
Turkey	2.6
Vietnam	2.5
China	2.4
Poland	2.4
Russia	2.3
Argentina	1.5

2.7.7 In Terms of Number of IP Addresses

The Spamhaus Project (which measures spam sources in terms of number of IP addresses used for spamming, rather than volume of spam sent) ranks the top three as the United States, China, and Russia, followed by Japan, Canada, and South Korea [14].

2.7.8 In Terms of Networks

As of 5 June 2007, the three networks hosting the most spammers are Verizon, AT&T, and VSNL International [13]. Verizon inherited many of these spam sources from its acquisition of MCI, specifically through the UUNet subsidiary of MCI, which Verizon subsequently renamed Verizon Business.

2.7.9 Most Common Products Advertised

According to information compiled by Table 3, e-mail spam can be broken down as follows [14].

Table 3 – E-mail Spam by Category

E-mail Spam by Category	
Products	25%
Financial	20%
Adult	19%
Scams	9%
Health	7%
Internet	7%
Leisure	6%
Spiritual	4%
Other	3%

2.8 How Spammers Operate

There are several steps in which spammers follows in order to prepare the spam e-mails to be sent to users and these steps are listed below:

2.8.1 Gathering of Addresses

In order to send spam, spammers need to obtain the e-mail addresses of the intended recipients and this is done through e-mail collection methods discussed above. Since spam is, by definition, unsolicited, this address harvesting is done without the consent (and sometimes against the expressed will) of the address owners. As a consequence, spammers' address lists are inaccurate. A single spam run may target tens of millions of possible addresses many of which are invalid, malformed, or undeliverable.

Sometimes, if the sent spam is "bounced" or sent back to the sender by various programs that eliminate spam, or if the recipient clicks on an unsubscribe link, that may cause that e-mail address to be marked as "valid", which is interpreted by the spammer as "send me more".

2.8.2 Obfuscating Message Content

Many spam-filtering techniques work by searching for patterns in the headers or bodies of messages. For instance, a user may decide that all e-mail they receive with the word "Viagra" in the subject line is spam, and instruct their mail program to automatically delete all such messages. To defeat such filters, the spammer may intentionally misspell commonly-filtered words or insert other characters, often in a style as in the following examples V1agra, Via'gra, Vi@graa, vi*gra, \/iagra. This also allows for many different ways to express a given work, making identifying them all more difficult for filter software. For example, using most common variations, it is possible to spell "Viagra" in over 1.3 * 1021 different ways.

The principle of this method is to leave the word readable to humans (who can easily recognize the intended word for such misspellings), but not likely to be recognized by a literal computer program. This is only somewhat effective, because modern filter patterns have been designed to recognize blacklisted terms in the various iterations of misspelling. Other filters target the actual obfuscation methods, such as the non-standard use of punctuation or numerals into unusual places. Similarly, HTML-based e-mail gives the spammer more tools to obfuscate text. Inserting HTML comments between letters can foil some filters, as can including text made invisible by setting the font color to white on a white background, or shrinking the font size to the smallest fine print. Another common trick involves presenting the text as an image, which is either sent along or loaded from a remote server.

As Bayesian filtering has become popular as a spam-filtering technique, spammers have started using methods to weaken it. To a rough approximation, Bayesian filters rely on word probabilities. If a message contains many words which are only used in spam, and few which are never used in spam, it is likely to be spam. To weaken Bayesian filters, some spammers now include lines of irrelevant, random words, in a technique known as Bayesian poisoning. A variant on this tactic may be borrowed from the Usenet abuser known as "Hipcrime" to include passages from books taken from Project Gutenberg, or nonsense sentences generated with "dissociated press" algorithms. Randomly generated phrases can create spoetry (spam poetry) or spam art.

Another method used to cover-up spam as legitimate messages is the use of auto generated sender names in the From field, ranging from realistic ones such as "Jackie F. Bird". Return addresses are also routinely auto-generated, often using unsuspecting domain owners' legitimate domain names, leading some users to blame the innocent domain owners. Blocking lists use IP addresses rather than sender domain names, as these are more accurate. A mail implying to be from example.com can be seen to be faked by looking for the originating IP address in the e-mail's headers. Sender Policy Framework, for example, helps by stating that a certain domain will only send e-mail from certain IP addresses.

Spam can also be hidden inside a fake "Undelivered mail notification" which looks like the failure notices sent by a mail transfer agent (a "MAILER-DAEMON") when it encounters an error [10].

2.8.3 Spam-support Services

A number of other online activities and business practices are considered by anti-spam activists to be connected to spamming. These are sometimes termed spam-support services like business services, other than the actual sending of spam itself, which permit the spammer to continue operating. Spam-support services can include processing orders for goods advertised in spam, hosting Web sites or DNS records referenced in spam messages, or a number of specific services as follows:

Some Internet hosting firms advertise bulk-friendly or bulletproof hosting. This means that, unlike most ISPs, they will not terminate a customer for spamming. These hosting firms operate as clients of larger ISPs, and many have eventually been taken offline by these larger ISPs as a result of complaints regarding spam activity. However, some spammers have managed to get what is called a pink contract, a contract with the ISP that allows them to spam without being disconnected.

A few companies produce spamware, or software designed for spammers. Spamware varies widely, but may include the ability to import thousands of addresses, to generate random addresses, to insert fraudulent headers into messages, to use dozens or hundreds of mail servers simultaneously, and to make use of open relays. The sale of spamware is illegal in eight U.S. states.

The alleged millions CDs are commonly advertised in spam. These are CD-ROMs supposedly containing lists of e-mail addresses, for use in sending spam to these addresses. Such lists are also sold directly online, but also often contain invalid addresses. In recent years, these have fallen almost entirely out of use due to the low quality e-mail addresses available on them, and because some e-mail lists exceed 20GB in size. The amount you can fit on a CD is no longer significant.

2.9 Spam E-mail Delivery

The spammers can use several ways in which they are able to deliver the spam to the users and the spammers use these ways or choose such kinds of ways in order to hide their identity and these ways are listed as follows

2.9.1 Webmail

A common practice of spammers is to create accounts on free webmail services, such as Hotmail, to send spam or to receive e-mailed responses from potential customers. Because of the amount of mail sent by spammers, they require several e-mail accounts, and use web bots to automate the creation of these accounts.

In an effort to cut down on this abuse, many of these services have adopted a system called the captcha in which the users attempting to create a new account are presented with a graphic of a word, which uses a strange font, on a difficult to read background. Humans are able to read these graphics, and are required to enter the word to complete the application for a new account, while computers are unable to get accurate readings of the words using standard OCR techniques. Blind users of captchas typically get an audio sample.

Spammers have, however, found a means of overcoming this measure. As it has been heard, they have set up sites offering free services and in order to get access to the site, a user displays a graphic from one of these web mail sites, and must enter the word. Once the bot has successfully created the account, the user gains access to the free service. Furthermore, standard image processing techniques work well against many captchas [8].

2.9.2 Third-party Computers

Recently, spammers discovered that if they sent large quantities of spam directly from their ISP accounts, recipients would complain and ISPs would shut their accounts down. Thus, one of the basic techniques of sending spam has become to send it from someone else's computer and network connection. By doing this, spammers protect themselves in several ways in which they hide their tracks, get others' systems to do most of the work of delivering messages, and direct the efforts of investigators towards the other systems rather than the spammers themselves. The increasing broadband usage gave rise to a great number of computers that are online as long as they are turned on, and whose owners do not always take steps to protect them from malware. A botnet consisting of several hundred compromised machines can effortlessly send out millions of messages per day, and would also complicate the tracing of spammers.

2.9.3 Open Relays

In the 1990s, the most common way spammers did this was to use open mail relays. An open relay is an MTA (Mail Transfer Agent), or mail server, which is configured to pass along messages sent to it from any location, to any recipient. In the original SMTP mail architecture, this was the default behavior where a user could send mail to practically any mail server, which would pass it along towards the intended recipient's mail server.

The standard was written in an era before spamming when there were few hosts on the internet, and those on the internet put up with a certain level of conduct. While this cooperative, open approach was useful in ensuring that mail was delivered, it was vulnerable to abuse by spammers. Spammers could forward bundles of spam through open relays, leaving it to the relays to deliver them.

In response, mail system administrators concerned about spam began to demand that other mail operators configure MTAs to cease being open relays. The first DNSBLs, such as MAPS RBL and the now inactive ORBS, aimed chiefly at allowing mail sites to refuse mail from known open relays. By 2003 less than 1% of corporate mail servers were available as open relays, down from 91% in 1997 [10].

2.9.4 Open Proxies

Within a few years, open relays became rare and spammers resorted to other strategies, most notably the use of open proxies. A proxy is a network service for making indirect connections to other network services. The client connects to the proxy and instructs it to connect to a server. The server attains an incoming connection from the proxy, not the original client. Proxies have many purposes, including Web-page caching, protection of privacy, filtering of Web content, and selectively bypassing firewalls.

An open proxy is one which will create connections for any client to any server, without authentication. Like open relays, open proxies were once relatively common, as many administrators did not see a need to restrict access to them.

A spammer can direct an open proxy to connect to a mail server, and send spam through it. The mail server logs a connection from the proxy not the spammer's own computer. This provides the spammer with a better cover up than an open relay, since most relays log the client address in the headers of messages they pass. Open proxies have also been used to hide the sources of attacks against other services besides mail, such as Web sites or IRC servers.

Besides relays and proxies, spammers have used other insecure services to send spam. One example is FormMail.pl, a CGI script to allow Website users to send e-mail feedback from an HTML form [9]. Several versions of this program, and others like it, allowed the user to redirect e-mail to random addresses. Spam sent through open FormMail scripts is frequently marked by the program's characteristic opening line "Below is the result of your feedback form."

As spam from proxies and other spammable resources grew, DNSBL operators started listing their IP addresses, as well as open relays. Today, spammers use infected client computers to deliver spam. Many still rely on Web hosting services of spam-friendly ISPs to make money.

2.9.5 Spammer Viruses

In 2003, spam investigators saw a major change in the way spammers sent spam. Rather than searching the global network for exploitable services such as open relays and proxies, spammers began creating services of their own. By installing computer viruses designed to deploy proxies and other spam-sending tools, spammers could harness hundreds of thousands of end-user computers. The widespread change from Windows 9x to Windows XP for many home computers, which started in early 2002 and was well under way by 2003, greatly accelerated the use of home computers to act as remotely controlled spam proxies. The original version of Windows XP as well as XP-SP1 had several major vulnerabilities that allowed the machines to be compromised over a network connection without requiring actions on the part of the user. While Windows 2000 had similar vulnerabilities, that operating system was never widely used on home computers.

Most of the major Windows e-mail viruses of 2003, including the Sobig and Mimail virus families, functioned as spammer viruses that are designed expressly to make infected computers available as spamming tools [9]. Besides sending spam, spammer viruses serve spammers in other ways. Beginning in July 2003, spammers started using some of these same viruses to carry out distributed denial-of-service (DDoS) attacks upon DNSBLs and other anti-spam resources. In August of that year, engineering company Osirusoft ceased providing DNSBL mirrors of the SPEWS and other block lists, after several days of unceasing attack from virus-infected hosts [10]. The very next month, DNSBL operator Monkeys.com submitted to the attacks as well. Other DNSBL operators, such as Spamhaus, have deployed global mirroring and other anti-DDoS methods to resist these attacks.

Zombie networks are particularly active in North America where about half of the Internet users are on a broadband connection and many leave their computers on all the time. In January, 2008, 8% of all e-mail spam was sent by the Storm botnet, created by the Storm Worm, first released in January, 2007. It is estimated that 1 million or more computers have been infected and their owners are unwilling and unknowing participants. In the 3rd quarter of 2008 almost one in every 400 e-mail messages contained a dangerous attachment, designed to infect the recipient's computer, eight times as often as in the previous quarter [5].

2.10 Authenticating E-mail Sender

Domain Authentication is an emerging issue, Since when an e-mail is sent you cannot be sure if it is coming from the address specified in the FROM, thus there is a big possibility that it is being sent from a spammer who is using the address in order to deliver his spam e-mails. Therefore, the need for authenticating the E-mails that are sent is becoming a necessity which can be a helpful factor in determining whether such an e-mail is a spam or not and contributes to the spam filtering efficiency.

3. Existing Solutions

This paragraph lists various solutions for tackling spam and image based spam, where the light is shed on the process and technique used to battle spam and the different features each solution contains. Also, the filtering steps that each solution requires to detect and prevent spam are presented.

3.1 Symantec

Symantec is considered one of the important firms that specialize in security products including anti-spam ones and below the Symantec's Bright mail anti-spam product along with its components and their features are discussed here in.

3.1.1 Spammers Employing Traditional Techniques

Security researchers at Symantec state that spammers have not discarded their old methods. Actually, in a wave of latest malware and spam crusades, spammers have revised and combined two oldest and commonly used topics. Symantec experts inform that they have observed the coming back of

spam mails which hide their malicious content in HTML code embedded in the form of mail attachments. It is a known obfuscation technique which has been discarded in favor of other methods such as image spam.

Symantec also reveals that the image spam, responsible for the major increase in spam activity during May 2009, became even more constant in June 2009, accounting for between 8% and 10% of the total spam detected by the security vendor. Actually, what they fear is that these spam attacks will probably follow ever more diverse strategies in times to come as spammers are collectively working to advance their attack vectors. Mayur Kulkarni, Researcher at Symantec, claims that spammers do not have to discover new methods to enter user's inbox. They can very well use the existing method with even better results, as reported by security watch week on July 7, 2009. Lastly, the security vendor has asked users that they should not carelessly open any attachments especially when it is sent by an unknown sender. With 419 spam mails, e-mail users are suggested not to reply fake appeals and do not show interest in any of the money making plans.

3.1.2 Symantec Brightmail AntiSpam

Symantec Brightmail AntiSpam™ offers complete, server-side anti-spam and antivirus protection. It actively seeks out, identifies, analyzes, and ultimately defuses spam and virus attacks before they trouble the users and overwhelm or damage the networks. Symantec Brightmail software that is installed at your site allows unwanted mail to be removed before it reaches the users' inboxes, without violating their privacy.

3.1.2.1 How Symantec Brightmail AntiSpam Works

Symantec Brightmail AntiSpam employs the following four major types of filters. First, AntiSpam Filters are created by Symantec using the state-of-the art technologies and strategies to filter and classify e-mail as it enters the site, Second, Content Filters are custom content filters are written by the user, using the Brightmail Control Center or the Sieve scripting language, to tailor filtering to the needs of the organization. Third, Allowed and Blocked Senders Lists in which lists can be created of allowed senders and blocked senders and third party lists can also be used. The lists included in the Brightmail Reputation Service are deployed by default. Fourth, Antivirus Filters in which Antivirus definitions and engines protect the users from e-mail borne viruses.

3.1.2.2 Features of Symantec Brightmail AntiSpam

AntiSpam Filtering Feature includes Heuristics that is a practical approach which targets patterns common in spam, Signatures that are Accurate and responsive approach that identifies the underlying "DNA" of evolving spam attacks. Defeats HTML-based and other evasion strategies used by spammers, Header that is similar to the Heuristics Filter, but applied to message headers, URL that matches the embedded URLs with a database of known spam URLs, Suspect List which Blocks e-mail from known spam senders (part of the Brightmail Reputation Service), Open Proxy List that blocks e-mail from insecure proxy servers by testing against the IP address of e-mail (part of the Brightmail Reputation Service), Safe List that allows e-mail from known clean domains (part of the Brightmail Reputation Service), Block and Allowed Senders Lists are Lists of trusted and blocked senders, IP connections, and domains created by administrators to augment Brightmail filtering, Content filters that are special purpose filters created by administrators to enforce organization-specific e-mail policies, and Third party filters which has easy integration with DNS-based blacklist and filtering services.

Other Filtering Features are group policies that specify groups of users, identified by e-mail addresses or domain names, and customize mail filtering for each group. Deployment options include gateway layer, internal relay layer, and e-mail server. The e-mail client add-ins for handling spam having Plug-ins for Outlook and Notes, and Web-based, with configurable notification option for recipients. Available antivirus protection detects and removes e-mail-borne viruses Quarantine Web-based, with configurable notification option for recipients. Spam management options in which to deliver the message normally, delete the message, deliver the message to the recipient's Spam folder, foldering agent moves spam to a designated folder in the end-user's mailbox, save the message to disk for administrator review, sends the message to an administrative account for further study, routes spam to a Web-based quarantine where recipients can review caught spam, and modify the message by adding configurable X-Header or subject line text to the message. Reporting and Statistics made up of standard interactive reports based on total spam or total virus messages found, and extended tracking and reporting of recipient, sender, domain, and other fields.

3.1.2.3 Symantec Brightmail AntiSpam Architecture

Symantec Brightmail AntiSpam consists of several components. The key components you need to consider are the following:

- Each Symantec Brightmail AntiSpam installation can have one or more Brightmail Scanners. Brightmail Scanners perform the actual filtering of e-mail messages.
- Each Brightmail Scanner contains a Brightmail Agent, and One or both of a Brightmail Server, and a Brightmail Client. If the Brightmail Scanner contains a Brightmail Client, then a supported mail transfer agent (MTA) must also reside on the same computer.
- The Brightmail Client is a communications channel between the MTA and the Brightmail Server. You can use multiple Brightmail Clients each one can talk to multiple Brightmail Servers. The Brightmail Client performs load balancing

between Brightmail Servers. The Brightmail Servers at your site process spam based on configuration options you select. Each Brightmail Server is a multi-threaded process that listens for requests from Brightmail Clients. Using a variety of state-of-the-art technologies, the Brightmail Server filters messages for classification. The classification, or verdict, is then returned to the Brightmail Client for successive delivery action.

- The Conduit connects to the BLOC to determine whether updated filtering rules are available. If new rules are available, the Conduit retrieves the updated rules using secure HTTPS file transfer. After authenticating the rules, the Conduit notifies the Brightmail Server to begin using the updated rules. The Conduit also manages statistics, both for use by the BLOC and in a local statistics pool for the generation of local reports. Each Symantec Brightmail AntiSpam installation has exactly one Brightmail Control Center. This is the central nervous system of your Symantec software. The Brightmail Control Center communicates with the Brightmail Agent on each of your Brightmail Scanners. For smaller installations, you can install the Brightmail Control Center and the Brightmail Scanner on the same computer. From this Web-based graphical user interface, you can configure start and stop each of your Brightmail Scanners, specify e-mail filtering options for groups of users or for all of your users at once, monitor consolidated reports and logs for all Brightmail Scanners, view summary and status information, administer Brightmail Quarantine, and view online help for Brightmail Control Center screens.

The Brightmail Control Center contains the following Features:

- Brightmail Quarantine provides storage of spam messages and Web-based end user access to spam. You can also configure Brightmail Quarantine for administrator-only access. Use of Brightmail Quarantine is optional.
- A single MySQL database stores all of your Symantec Brightmail AntiSpam configuration information, as well as Brightmail Quarantine information and e-mails (if you are using Brightmail Quarantine). Configuration information is communicated to each Brightmail Scanner via an XML file. A Java-based Web Server (by default this is the Tomcat Web Server) performs Web hosting functions for the Brightmail Control Center and Brightmail Quarantine.

3.1.2.4 Symantec Brightmail AntiSpam Filtering Process

With the default configuration, the filtering process works as follows. First, the SMTP server receives the mail message and processes any security settings. Second, the Brightmail Client (integrated with the MTA) sends a copy of the mail message to the Brightmail Server. Third, by default the Brightmail Server processes mail in the following order, allowed senders you identify, blocked senders you identify, Symantec Brightmail AntiSpam filters, content filters you create and finally the Brightmail Server returns the verdict of the message to the Brightmail Client. Fourth, the Brightmail Client tells the SMTP server to perform the appropriate action, based on the policies in place [15]. The Bright mail anti-spam solution is composed of several components and these components need to interact with each other in order to provide the feature that is needed from it and these interactions are shown in Figure 1.

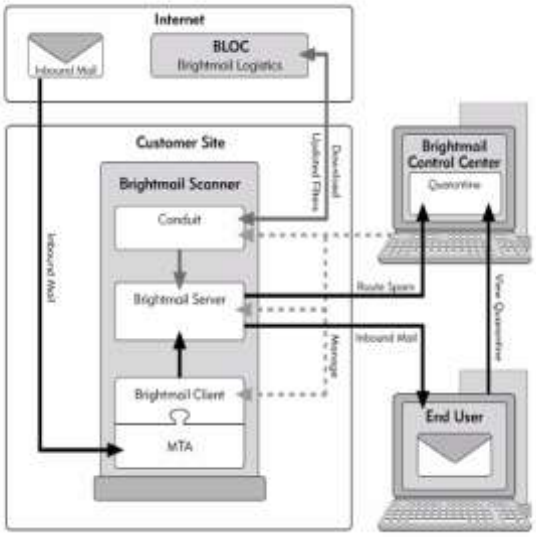

Fig-1: Symantec Brightmail Components Interaction

3.2 Kaspersky

Although Internet Security suites usually include as standard e-mail spam filter, spammers continue to find ways around the checks that are made. One of those workarounds is the use of images with text buried in the image data. This kind of spam can be checked for, but currently it is done using machine recognition. Spammers can overcome those checks by making the text fuzzy and adding distortion or rotation to an image. Kaspersky Lab has a statistics-based method for detecting image-based spam that is used to bypass traditional text-based filters. The technology analyses whether text is contained in images based on the graphic pattern of words and lines, said developer Eugene Smirnov. Spam is expected to continue to be a problem in 2009, particularly with the rise in the number and popularity of websites that allow user-generated content.

Kaspersky Anti-Spam 3.0 provides thorough and accurate protection from spam for users of corporate mail systems and public e-mail services.

3.2.1 Kaspersky Anti-Spam

There are several features that are offered by Kaspersky Anti-spam solution and these features include the following.

3.2.1.1 Protection from Spam

List-based filtrations in which sender's IP addresses are checked against blacklists of spammers, which are maintained by Internet service providers and public organizations (DNS-based Blackhole Lists). System administrators can add addresses of trusted correspondents to a safe list, ensuring that their messages are always delivered without undergoing filtration. Analysis of formal attributes where the program recognizes spam by such typical characteristics as distorted sender addresses or the absence of the sender's IP address in DNS, an excessive number of intended recipients or hidden addresses. The size and format of messages are also taken into consideration. Linguistic heuristics where the program scans messages for words and phrases that are typical of spam messages. Both the content of the message itself and any attachments are analyzed. Graphic spam in which a database of signatures for graphic spam equips the program to block messages containing spam images, a type of spam that has become increasingly common in recent years. Real-time UDS requests where the Urgent Detection System is updated with information on spam messages literally seconds after they first appear on the Internet. Messages that could not be assigned a definitive status (e.g., spam, no-spam) can be scanned using UDS. The e-mail that is received passes into a process of message analysis as shown in Figure 2 and includes several analysis procedures in order to analyze the message

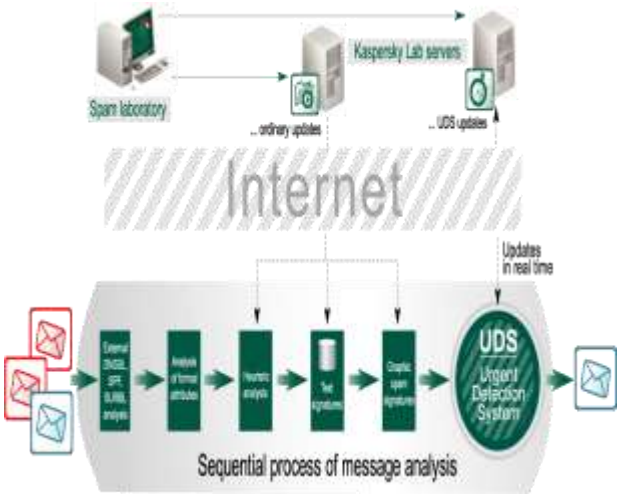

Fig-2: Kaspersky E-mail analysis process

3.2.1.2 Administration

Flexible management in which the web interface allows system administrators to manage the application both locally and remotely. The filtration level is easily configurable, as are blacklists and safe lists. It is also possible to disable/enable individual filtration rules. Management of user groups where the administrator can create user groups either using lists of addresses or domain masks (for example, XXX@domain.com) and apply individual settings and filtration rules to each group. Options for processing spam where the program can be configured to process spam by either automatically deleting it, redirecting it to the quarantine folder with a note to the user or sent for further filtration to the mail client. Detailed reports where the administrators can easily monitor the application, the protection status and license status, using HTML reports or alternatively, by viewing log files. Data can be exported in CSV and Excel formats.

3.2.2 Kaspersky Anti-Spam 3.0 MP1 Critical Fix

The following improvements have been introduced since Kaspersky Anti-Spam 3.0 MP1 (3.0.255.0) where methods for fighting so-called "graphic" spam, i.e. tools used to analyze graphic attachments. New algorithms have been introduced for processing and identification of similar images with textual content as well as the GSG-8 and GSG-9 technologies. The following problems have been fixed as compared to Kaspersky Anti-Spam 3.0 MP1 CF1 (3.0.274.0) where possible termination or freezing of filtering processes when a list of protected domains is used, and accidental setting of incorrect access rights for the files of application components if they were previously updated using a package of modified application files from previous product versions [16].

3.3 Trend micro

It's no longer efficient to compile lists of known spammers and filter them out, because those lists are so large and growing bigger all the time, adds Hemmendinger. And it's too cumbersome to update them on a daily basis. ''What we've learned over time is the more commonly used methods would be content filtering, like text filters that look for certain key words or sophisticated heuristics that look at the content of a message to see if it appears to fit the mold of what is readily recognized as spam,'' Hemmendinger says. He also pointed to techniques that spammers use to trip up e-mail filters, like adding asterisks between each letter in a word so it can't be identified.'' With the release of InterScan Messaging Server Suite (IMSS), Trend Micro strives to provide solution providers with effective tools to battle spam and protect users from increasing ills associated with e-mail, ranging from script bombs to worm-bearing messages.

In Figure 3 we can have a view on a snap shot of the Trend Micro IMS anti-spam solution which shows the configuration that can be altered or given by the user

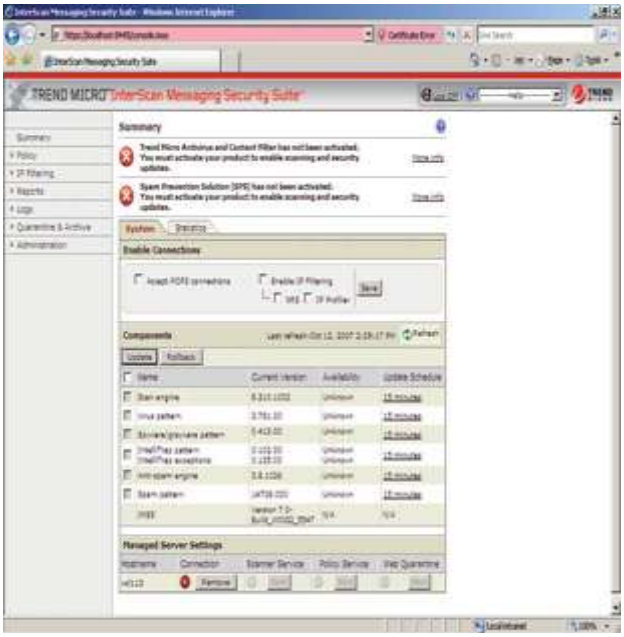

Fig-3 : Trend Micro InterScan Messaging Security

3.3.1 InterScan Messaging Server Suite Filtering Process

First, a message passes through Trend Micro's 32-bit virus scan engine. After the messages are checked for viruses, they're passed off to the content management portion of IMSS, that process is the key to battling spam and other e-mail-related problems. Trend Micro directs the advantage of policies toward content filtering, and those policies allow complete control of e-mail beyond spam management. Solution providers can script policies that prevent confidential data from being transmitted or create policies that identify unwanted messages. Policies clearly define what acceptable use of company e-mail is and what is not.

The primary reason for using IMSS is controlling spam. While policies can offer some protection from spam, the real answer to effectively fighting it lies with automation. IMSS employs complex heuristics to identify spam. Every message is examined for phrases or content that fits the profile of a spam message, and anti-spam heuristics can be tuned to filter based on content and determine how aggressively the antispam filtering should be applied. Administrators have several options for handling e-mail identified as spam. They can add the word "spam" to the subject line, redirect the suspect e-mail or quarantine the e-mail.

3.3.2 Trend Micro Spam Prevention Solution

Spam Prevention Solution offers a comprehensive, multi-tiered spam and phishing defense. Three distinct tiers of anti-spam protection include E-mail Reputation, IP Profiler, and the anti-spam composite engine. The solution uses multiple techniques to keep threats completely off of the network, securing the network and preserving bandwidth, storage, and other network resources. Spam Prevention Solution includes patent-pending image spam detection technology and other cutting-edge approaches to protect organizations as spam and phishing threats evolve where it blocks most spam before it even reaches the gateway, uses the world's largest most trusted reputation database, deploy dynamic reputation services to stop zombies and botnets as they first emerge, blocks e-mail senders that exceed threat thresholds set by the organization providing protection customized to the organization's e-mail traffic, delivers automatic customer specific reputation services to stop spam, creates a firewall against bounced mail attacks, and combines multiple protective techniques including statistical analysis, advanced heuristics, whitelists, and blacklists. Also, it includes Features image spam detection and other cutting-edge technologies, content filtering and expanded language support to improve spam protection for global companies, provides dedicated anti-phishing techniques, including signatures, and reputation services to stop both corporate and consumer phishing attacks. Furthermore, it offers single Web-based management console to customize spam tolerance settings, create approved sender lists, establish filter actions, and set policies for individuals or groups. Moreover, it simplifies administration through LDAP integration, delegated administration, and message tracking. In addition to enabling end users to manage their own spam with Web-based End-User Quarantine and quarantine notification e-mails.BENEFITS

3.3.3 Policies or Rule Based Detection Mishaps

Antivirus firm Trend Micro unwittingly targeted the letter "P" with a recent rules update, forcing all e-mail containing the objectionable letter into quarantine. According to their knowledge base article titled Solution 14638, "Antispam Rule 915 unintentionally blocks some legitimate e-mails scanned by InterScan eManager and ScanMail eManager." The cause is the letter P.

According to Trend Micro, the problem affects their Internet gateway, e-mail and groupware products, including InterScan Messaging Security Suite, InterScan eManager, ScanMail for Exchange, ScanMail eManager, and ScanMail for Lotus Notes. A spokesman for Trend Micro declined to comment on the issue, stating only that "we've notified customers and resellers." According to Internet Week, much of that contact was done via e-mail. One can only imagine the difficulty of composing an e-mail describing the nature of the problem while simultaneously avoiding the use of the letter P.

Trend Micro advises that the unfortunate P mishap can be resolved by updating to Antispam Rule 916 or later. Several of their products include options to resend e-mails erroneously quarantined by the filtering rules. Their Knowledge Base article Solution 14638 contains links to the support solutions for these products [17].

3.4 Mail-Secure

The Mail-Secure anti-spam solution is a product of the PineApp firm which uses pattern detection and includes the following features.

3.4.1 Image Spam Defense

Spammers are consistently creating sophisticated new weapons in their arms race with anti-spam technology, the latest of which is image-based spam. The number of unsolicited messages containing images has grown significantly throughout 2006, and is expected to continue to grow and spread.

Through constant monitoring, PineApp has identified that image-based spam tends to be distributed in massive waves at one of the distribution peaks, PineApp measured image-based spam as 30% of all global spam. Image-based spam creates bandwidth and storage problems, since the typical image based spam message weighs more than three times that of a regular spam message. At the image-spam distribution peaks, the bandwidth and storage requirements increase upwards of 70%. Also, Image-based spam is a new and growing problem leading to loss of productivity and a drain on IT resources, most anti-spam solutions have problems dealing with image-based spam, and by dealing with it ineffectively they create other problems along the Way. Thus, PineApp has implemented a unique solution to decode images, and treat them with RPD similarly to other types of spam which improves the already superior spam catch rate, and maintains low false positive rate

3.4.2 Newest Trends in Image-Based Spam

Lately, spammers have been experimenting with new techniques such as broken images i.e. splitting a single image into smaller images that fit together like puzzle pieces. This technique makes it even more difficult for anti-spam engines to catch and block.

3.4.3 Mail-SeCure Filtering Process

The web-based interface, presented to the user upon logging in, is very easy to use and clutter free. The interface presents its data in a clear and straightforward manner, with minimal delay when saving any configuration changes. For added security, the interface also includes a timeout function, returning the administrator to the login page after a set period of inactivity. Mail-SeCure's method of protecting against spam is controlled through the use of policies.

Mail-SeCure is a leading perimeter security appliance that protects all sized organizations (from 50 up to 10,000 users), from both targeted and non-targeted e-mail-related threats such as spam, viruses and malicious code. Mail-SeCure from PineApp is a gateway level device designed to offer e-mail protection to small or medium sized companies with support for up to 500 users. While this test was primarily concerned with spam detection, it should be noted that Mail-SeCure also provides protection from e-mail borne malware.

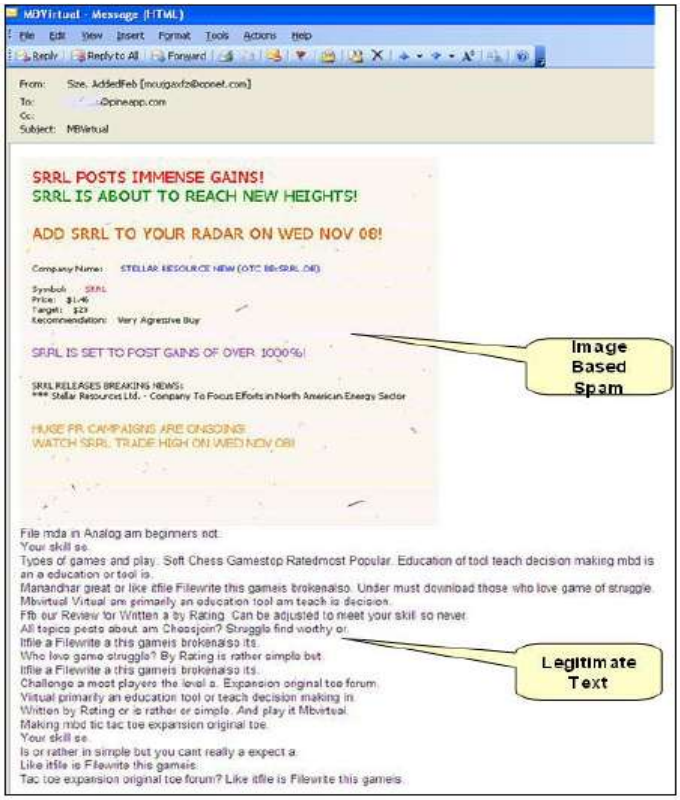

Fig- 4: E-mail Spam sample

Configuration of Mail-SeCure is made simple by the provision of a well-written and easy to follow quick installation guide. Within the policies configuration screen, there are four separate rule groups available to the administrator. These are Attachment, Spam, General, and Black & White Rules. Each of these rule groups shares a similar layout, allowing for familiarization with the method by which these rules may be configured. When dealing with spam, Mail-SeCure splits the traffic into one of three types Local to Local, Remote to Local, and Local to Remote, effectively covering both internal and external mail. Each of these three traffic types may have its own policy. For the purposes of reviewing statistics relating to processed e-mail, Mail-SeCure provide five separate report pages. Included among these are Summary, Reports, User Reports, Domain Reports, and Statistics [18] [19].

Figure 4 shows a sample e-mail message that contains two parts within its body, one part is an image that has written spam text embedded in it and the other is a legitimate text written beneath the image crafted in order to foil anti-spam solutions.

3.5 Publications and Literature

Zhe Wang, William Josephson, Qin Lv, Moses Charikar, Kai Li[20] in Filtering Image Spam with Near-Duplicate Detection propose an image spam detection system that uses near-duplicate detection to detect spam images, they rely on traditional anti-spam methods to detect a subset of spam images and then use multiple image spam filters to detect all the spam images that "look" like the spam caught by

traditional methods. Battista Biggio, Giorgio Fumera, Ignazio Pillai, Fabio Roli[21] in Image Spam Filtering by Content Obscuring Detection propose an approach based on low-level image processing techniques to detect one of the main characteristics of most image spam, namely the use of content obscuring techniques to defeat OCR tools by finding the noise level of a certain image spam. Jason R. Bowling, Priscilla Hope, Kathy J. Liszka[22] in Spam Image Identification Using an Artificial Neural Network propose a method for identifying image spam by using FANN (Fast Artificial Neural Network) library model and training the artificial neural network. A detailed process for preprocessing spam image files is given, followed by a description on how to train an artificial neural network to distinguish between ham and spam. M. Muztaba Fuad, Debzani Deb, M. Shahriar Hossain[23] in A Trainable Fuzzy Spam Detection System presents the design and implementation of a trainable fuzzy logic based e-mail classification system that learns the most effective fuzzy rules during the training phase and then applies the fuzzy control model to classify unseen messages. M. Soranamageswari, C. Meena[24] in Statistical Feature Extraction for Classification of Image Spam Using Artificial Neural Networks present an experimental system for the classification of image spam by considering statistical image feature histogram and mean value of an block of image. A comparative study of image classification based on color histogram and mean value is presented.

Ms.D.Karthika Renuka, Dr.T.Hamsapriya, Mr.M.Raja Chakkaravarthi and Ms.P.Lakshmisurya (2011) performed a comparative analysis on spam classification based on supervised learning using several machine learning techniques. In this analysis, the comparison was done using three different machine learning classification algorithms viz. Naïve Bayes, J48 and Multilayer perceptron (MLP) classifier. Results demonstrated high accuracy for MLP but high time consumption. While Naïve Bayes accuracy was low than MLP but was fast enough in execution and learning. The accuracy of Naïve Bayes was enhanced using FBL feature selection and used filtered Bayesian Learning with Naïve Bayes. The modified Naïve Bayes showed the accuracy of 91% as in [25].

Rushdi Shams and Robert E. Mercer (2013) performed a comparative analysis on classification of spam emails by using text and readability features. This paper proposed an efficient spam classification method along with feature selection using content of emails and readability. This paper used four datasets such as CSDMC2010, Spam Assassin, Ling Spam, and Enron-spam. Features are categorized into three categories i.e. traditional features, test features and readability features. The proposed approach is able to classify emails of any language because the features are kept independent of the languages. This paper used five classification based algorithm for spam detection viz. Random Forest (RF), Bagging, Adaboostm 1, Support Vector Machine (SVM) and Naïve Bayes (NB). Results comparison among different classifiers predicted Bagging algorithm to be the best for spam detection as in [26]. Megha Rathi and Vikas Pareek(2013) performed an analysis on spam email detection through Data Mining by performing analysis on classifiers by selecting and without selecting the features as in [27].

Anirudh Harisinghaney, Aman Dixit, Saurabh Gupta and Anuja Arora (2014) performed a comparative analysis on text and images by using KNN, Naïve Bayes and Reverse-DBSCAN Algorithm for email spam detection. This analysis paper proposed a methodology for detecting text and spam emails. They used Naïve Bayes, K-NN and a modified Reverse DBSCAN (Density- Based Spatial Clustering of Application with Noise) algorithm's. Authors used Enron dataset for text and image spam classification. They used Google's open source library, Tasseract for extracting words from images. Results show that these three machine learning algorithms gives better results without preprocessing among which Naïve Bayes algorithm is highly accurate than other algorithms as in [28].

Savita Pundalik Teli and Santosh Kumar Biradar (2014) performed an analysis on effective email classification for spam and non-spam emails as in [29]. Izzat Alsmadi and Ikdam Alhami (2015) performed an analysis on clustering and classification of email contents for the detection of spam. This paper collected a large dataset of personal emails for the spam detection of emails based on folder and subject classification. Supervised approach viz. classification along-side unsupervised approach viz. clustering was performed on the personal dataset. This paper used SVM classification algorithm for classifying the data obtained from K-means clustering algorithm. This paper performed three types of classification viz. without removing stop words, removing stop words and using Ngram based classification. The results clearly illustrated that N-gram based classification for spam detection is the best approach for large and Bi-language text as in [30].

Ali Shafigh Aski and Navid Khalilzadeh Sourati (2016) performed an analysis using Machine Learning". This paper utilized three machine learning algorithms viz. Multi-Layer Neural Network, J48 and Naïve Bayes Classifier for detection of spam mails from ham mails using 23 rules. The model demonstrated high accuracy in case of MLP with high time for execution while Naïve Bayes showed slightly less accuracy than MLP and also low execution time as in [31].

4. Conclusion

Image Spam detection have been causing problems from the first day it was known and up till now with all the solutions that have been developed by various venders and users, it still poses a great threat and still able to penetrate to the user's e-mail and up till now various venders still look at enhancing and updating their algorithms in order to achieve a higher detection rate with lowers false positive, and the reason that keeps this ongoing problem is the ways that the spammers are employing to fool those algorithms. In this paper, we introduced some of the available solutions for tackling spam and image based spam, where the light is shed on the process and technique used to battle spam and the different features each solution contains.

REFERENCES

1. Cormack, Gordon V (2008), Email Spam Filtering: A Systematic Review. Now Publishers Inc, ISBN 978-1601981462
2. Gyöngyi, Zoltán; Garcia-Molina, Hector (2005), "Web spam taxonomy", Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb), 2005 in The 14th International World Wide Web Conference (WWW 2005) May 10, (Tue)-14 (Sat), 2005, Nippon Convention Center (Makuhari Messe), Chiba, Japan., N.Y.: ACM Press, ISBN 1-59593-046-9
3. Gonzalez, Rafael C.; Woods, Richard E. (2008), Digital Image Processing, 3rd edition, Prentice Hall, ISBN 9780131687288
4. 6 Stone, Brad (March 2009), Spam back to 94% of all Email New York Times, http://bits.blogs.nytimes.com/2009/03/31/spam-back-to-94-of-all-e-mail/
5. Schryen, Guido (2007), Anti-Spam Measures: Analysis and Design, Springer, ISBN 978-3540717485
6. Cormack, Gordon V (2008), Email Spam Filtering: A Systematic Review. Now Publishers Inc, ISBN 978-1601981462
7. Schwartz, Alan; Garfinkel, Simson (1998), Stopping Spam, O'Reilly, ISBN: 156592388X
8. Spammer-X; Posluns, Jeffrey; Sjouwerman, Stu (2004), Inside the SPAM Cartel, Syngress, ISBN 978-1932266863
9. Goodman, Danny (2004), Spam Wars - Our Last Best Chance to Defeat Spammers, Scammers, and Hackers, SelectBooks, Inc., ISBN 1-59079-063-4
10. Miller, Frederic P.; Vandome, Agnes F.; McBrewster, John (2009), E-mail: E-mail Spoofing, E-mail Privacy, HTML E-mail, E-mail Letter, Privacy-Enhanced Electronic Mail, Push E-mail, Anti-Spam Techniques, E-mail Spam, E-mail Address, VDM Publishing House, ISBN 978-6130066833
11. Sophos Plc (2008-07-15). "Only one in 28 e-mails legitimate, Sophos report reveals rising tide of spam in April - June 2008". http://www.sophos.com/pressoffice/news/articles/2008/07/
12. Brasil assume a liderança do spam mundial em 2009, diz Cisco (Portuguese), December 8, 2009, http://idgnow.uol.com.br/seguranca/2009/12/08/brasil-assume-a-lideranca-do-spam-mundial-em-2009-diz-cisco/
13. "Spamhaus Statistics: The Top 10". Spamhaus Blocklist (SBL) database. The Spamhaus Project Ltd. dynamic report, http://www.spamhaus.org/statistics/countries.lasso
14. Evett, Don (2006), "Spam Statistics 2006". http://spam-filter-review.toptenreviews.com/spam-statistics.html
15. Symantec Corporation (2008), Symantec Brightmail AntiSpam Deployment Planning Guide
16. Kaspersky Lab (2008), Kaspersky Anti-Spam 3.0 Administrators Guide
17. Trend Micro Incorperated (2008), Trend Micro ScanMail, InterScan Security Guide
18. PineApp Ltd (2007), Mail-SeCure Perimeter Security white paper
19. PineApp Ltd (August 2009), Mail-SeCure Image-Based Spam Treatment white paper
20. Zhe Wang; William Josephson; Qin Lv; Moses Charikar; Kai Li (2008), Filtering Image Spam with Near-Duplicate Detection, In Conference on E-mail and Anti-Spam (CEAS)
21. Battista Biggio; Giorgio Fumera; Ignazio Pillai; Fabio Roli (2008), Image Spam Filtering by Content Obscuring Detection, In International Conference on Image Analysis and Processing (ICIAP)
22. Jason R. Bowling; Priscilla Hope; Kathy J. Liszka (2009), Spam Image Identification Using an Artificial Neural Network, In International Conference on Artificial Neural Networks (ICANN)
23. M. Muztaba Fuad; Debzani Deb; M. Shahriar Hossain (2005), A Trainable Fuzzy Spam Detection System, In International Conference on Computer and Information Technology (ICCIT)
24. M. Soranamageswari; C. Meena (2010), Statistical Feature Extraction for Classification of Image Spam Using Artificial Neural Networks, In International Conference on Machine Learning and Cybernetics (ICMLC)
25. D. K. Renuka, T. Hamsapriya, M. R. Chakkaravarthi and P. L. Surya, "Spam Classification Based onSupervised Learning Using Machine Learning Techniques", in proc. IEEE- International Conference on Process Automation, Control and Computing, 2011, pp. 1–7.
26. R. Shams and R. E. Mercer, "Classifying spam emails using text and readability features", in proc. IEEE International Conference on Data Mining (ICDM), 2013, pp. 657–666..
27. M. Rathi and V. Pareek, "Spam Email Detection through Data Mining-A Comparative Performance Analysis", in International Journal of Modern Education and Computer Science, vol. 12, pp. 31-39, 2013.
28. A. Harisinghaney, A. Dixit, S. Gupta, and Anuja Arora, "Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN Algorithm", in proc. IEEE-International Conference on Reliability, Optimization and Information Technology (ICROIT), 2014, pp.153-155.
29. S. P. Teli and S. K. Biradar, "Effective Email Classification for Spam and Non- spam", in International Journal of Advanced Research in Computer and software Engineering, Vol. 4, 2014.
30. Alsmadi and I. Alhami, "Clustering and classification of email contents", in Journal of King Saud University - Computer and Information Science -Elsevier, vol. 27, no. 1, pp. 46–57, 2015.
31. A. S. Aski and N. K. Sourati, "Proposed efficient algorithm to filter spam using machine learning techniques", in Pacific Science Review- A Natural Science Engineering- Elsevier, Vol. 18, No. 2, pp. 145–149, 2016.