

A Review of Spam Mail Genre and Spam Filtering Techniques

Oluwatunde, S.J., Gbolagade, K.A., & Isiaka, R.M

Department of Computer Science

Kwara State University

Malete, Kwara State, Nigeria

E-mail: oluwatundesola@yahoo.com; kazeem.gbolagade@kwasu.edu.ng; abdulrafiu.isiaka@kwasu.edu.ng

Phone: +447909365912; +2348109668798; +2348035196608

ABSTRACT

Since the Internet started, its usefulness has also grown alongside the technological development. Basically, the Internet was created for information purposes which include; education, entertainment, communication (emails), etc. However, the more useful the Internet gets, the more end users of the email boxes have this problem of Spam. Spam has annoyed, confused and frustrated email users and their administrators, because the major cost of spam is borne by the recipient and their mail servers. These costs are in the terms of bandwidth wastage, storage space that the spam emails utilize and most importantly, the likelihood that very important email/information could be lost as anti-spam filters may wrongfully filter them out. Sometimes, the owner of the mailbox may get so overwhelmed by the volume of spam and deleting useful email. It is obvious that end users are no longer inundated with the same volume of unwanted email that they faced a few years ago, however spam is still a major problem for network operators because spam still accounts for around 70 percent of all Internet traffic, in spite of the numerous regulatory initiatives enacted throughout the world designed to combat the problem. In this paper, we review spamming techniques as it relates to different genre of spam mails and spam filers

Keywords: Spam, Security, E-Mail, Filtering, Genre, Users, Scammers, Spammers, Internet Traffic

CISDI Journal Reference Format

Oluwatunde, S.J. (2025): A Review of Spam Mail Genre and Spam Filtering Techniques. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 16 No 2, Pp 23-45.

Available online at www.isteams.net/cisdijournal. dx.doi.org/10.22624/AIMS/CISDI/V16N2P3

1. INTRODUCTION

All email messages are sent with an envelope address specifying the senders and the receivers email addresses. However, spammers fake the sending email address to evade being caught by some anti-spam filters. Figure 3 shows a list of emails where I have marked those that were purportedly sent by me. This faking of addresses by spammers is possible because the SMTP standard has not defined authentication of the sender e-mail address. SMTP only authenticates the receivers email address. Almost all of spam is produced by zombie computers of open relay websites. These are vast networks of personal computers that have been infected with malicious software, permitting spammers to use their computing power without the owners' knowledge. It is known that zombies now deliver about half of all spam.

Spam Definition

There is generally no single perfect definition for Spam, because there is a rough consensus as to what spam really is. One problem is that what I consider as spam may not be considered as spam by another user. However, users of email immediately recognize it as soon as it gets into their mailboxes. Spam comes in a variety of ways and format. Some are to advertise or to sell products while others aim at transmitting viruses to the computer system of the recipient. What is common to all the various definitions of spam is that they are unsolicited, not 'unwanted' emails and the intention of the spammer is to flood people's computers with information. Note however that an unsolicited email may actually be needed/ wanted by the receiver. It is good to note that the very first spam distributed on a wide area network dates as far back as 1978. This was an advertisement from a digital company. Until a perfect definition for spam is found, we are stuck with the following types/ definition of spam, presented below.

Unsolicited Bulk Email (UBE)

This refers to unsolicited emails sent in large quantities from a single source. These emails are sent to a large number of recipients and some of the email addresses do not even exist.

Unsolicited Commercial E-mail (UCE)

This refers to emails sent for the purpose of advertising products and other types of solicitations.

Fraudulent Email

These are emails sent to one or few people with the intention of defrauding them or trying to trick recipients into giving up bank and credit card information. It includes the popular Nigerian 419 emails.

Unsolicited Automated Email

There are people of the opinion that spam should be called unsolicited automated email. This definition includes some email that many legal definitions of spam don't. Legal definitions of spam tend to exclude mail sent by companies that have an "existing relationship" with the recipient. For example, buying an item from a company does not imply that you have solicited ongoing email from them. If they then send a stream of product sales emails, it is still spam.

They also argue that spam is not unsolicited commercial email. For example, if someone heard that I was looking for a car to buy and sent me an email offering to sell me one, I'd be delighted, and yet this email would be both commercial and unsolicited. It is merely incidental, too, that spam is usually commercial. If someone started sending mass email to support some political cause, for example, it would be just as much spam as email promoting a porn site.

Others

There are other types of spam that include emails messages sent where the sender's identity is forged or the message is sent through unprotected SMTP servers, unauthorized proxies or botnets. Fig 2 below shows a spam email with a forged senders address. Some spam messages actually arrive the mailbox blank, that is with nothing in the body of the message.

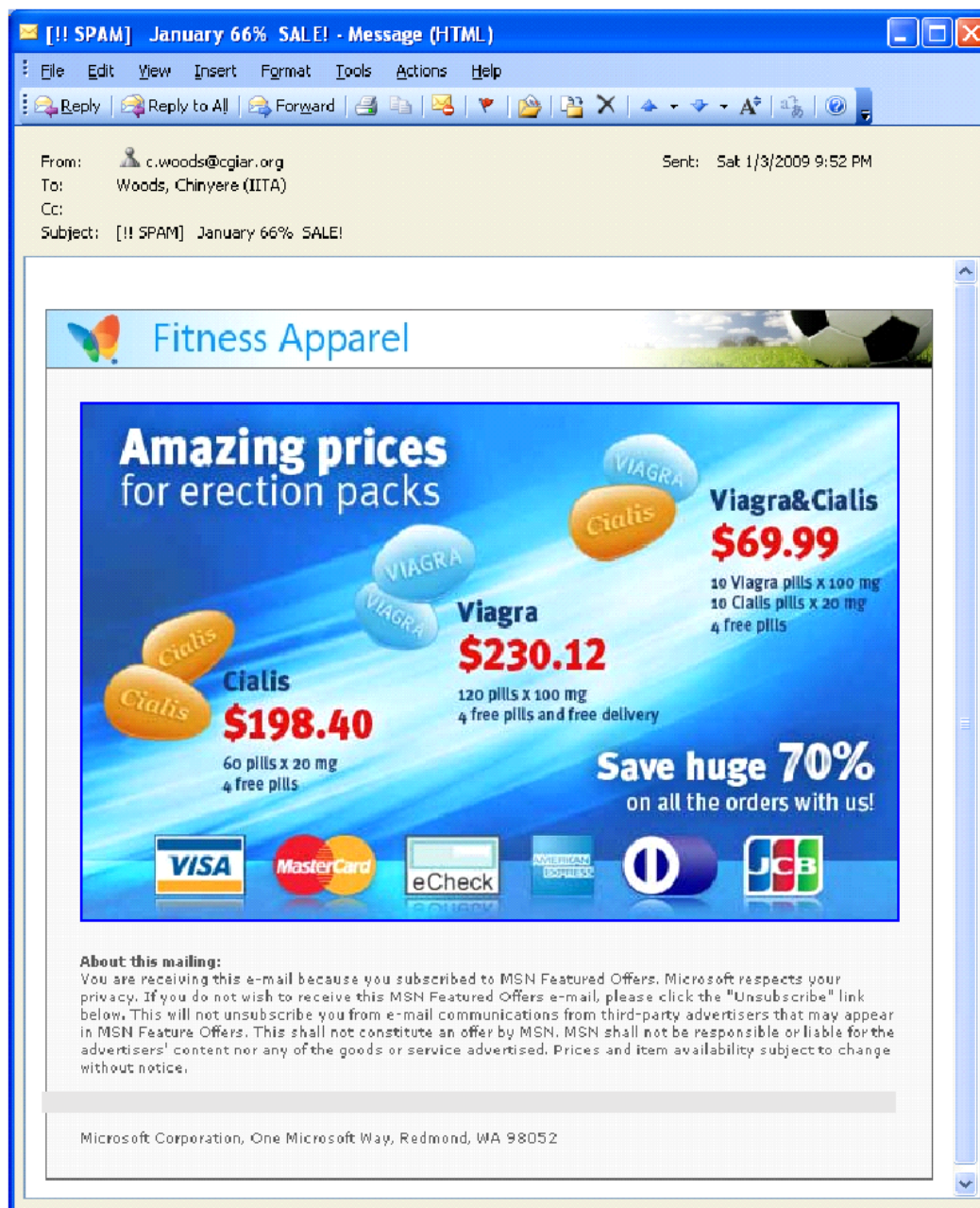


Figure 1: Email spam with a forged sender's identity

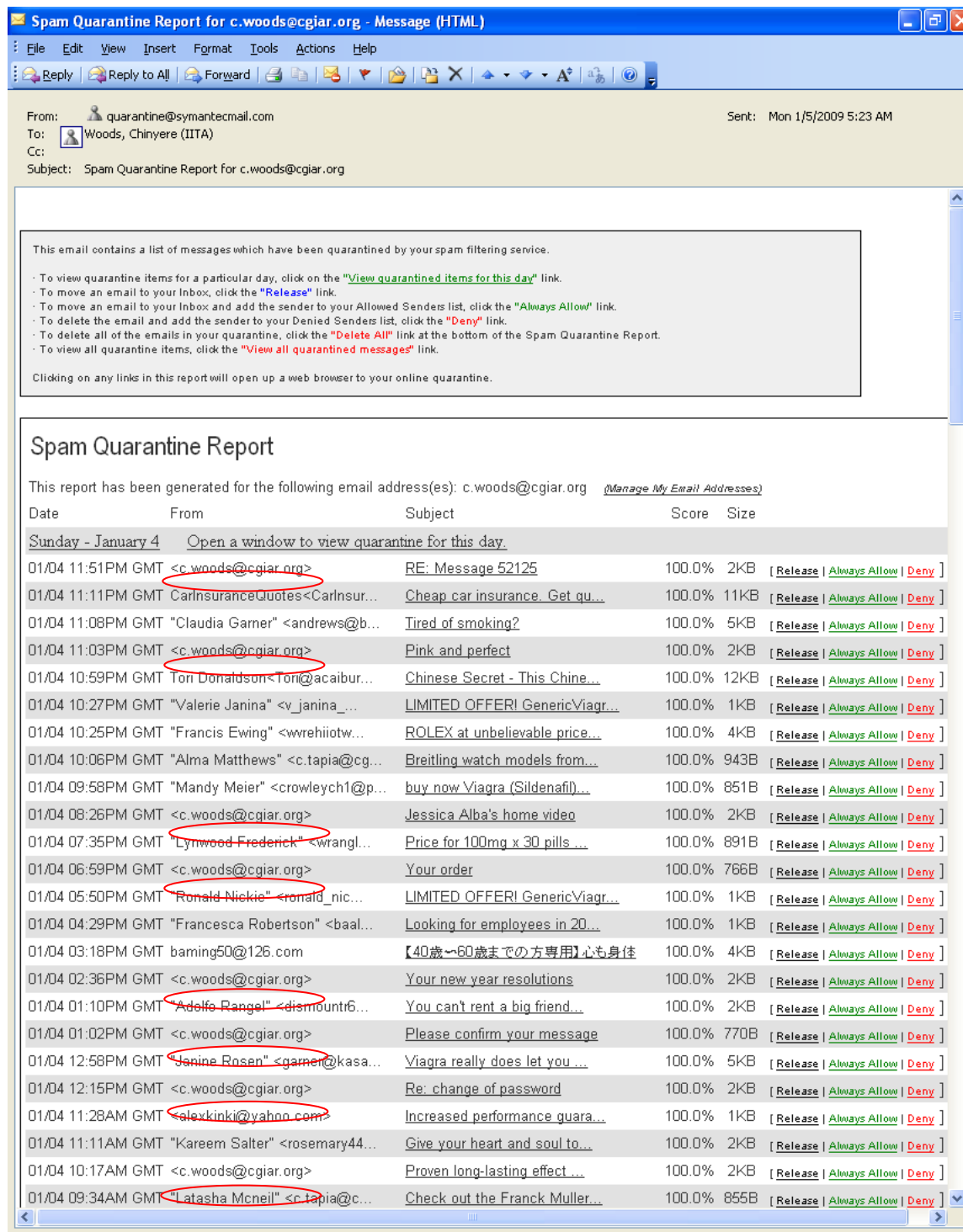


Figure 2: A list of spam emails caught by an anti-spam filter indicating emails that had forged sender's identification.

2. SPAM GENRE

Text-based spam:

Text-based email spams are spam emails that are sent where the content is only text. These used to be very common in the past but are now being contained because they are easier to filter. As email filtering technologies have become more sophisticated, bulk e-mailers have begun sending larger, image-based messages in an attempt to slip past anti-spam filters.

Image-based spam:

These are spam email messages sent where the whole body of the message is an image. Figure 4 shows some examples. The image could be graphic or photographic. The aim of spammers in resorting to this type of message is to beat the text-based spam filters. Sometimes these images are actually text converted or embedded in an image as shown in figure 4a. At other times, the images are split into tiny bits and when the pictures are fully downloaded, they all appear as a single image. Figure 5 illustrates this.

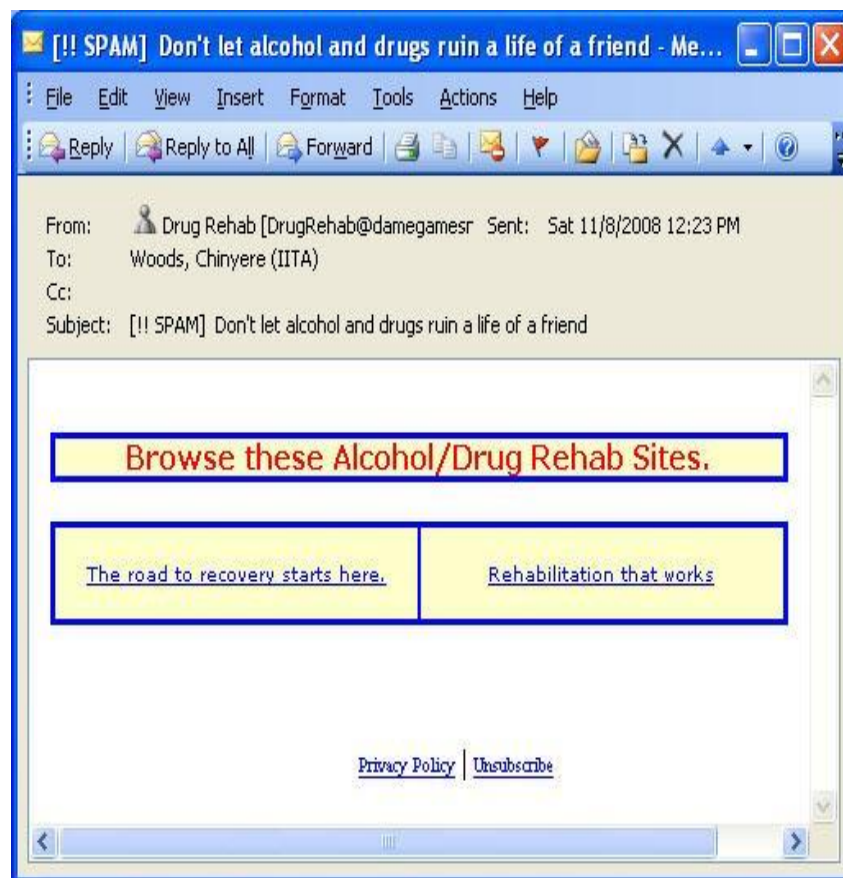


Figure 3: Text as image spam

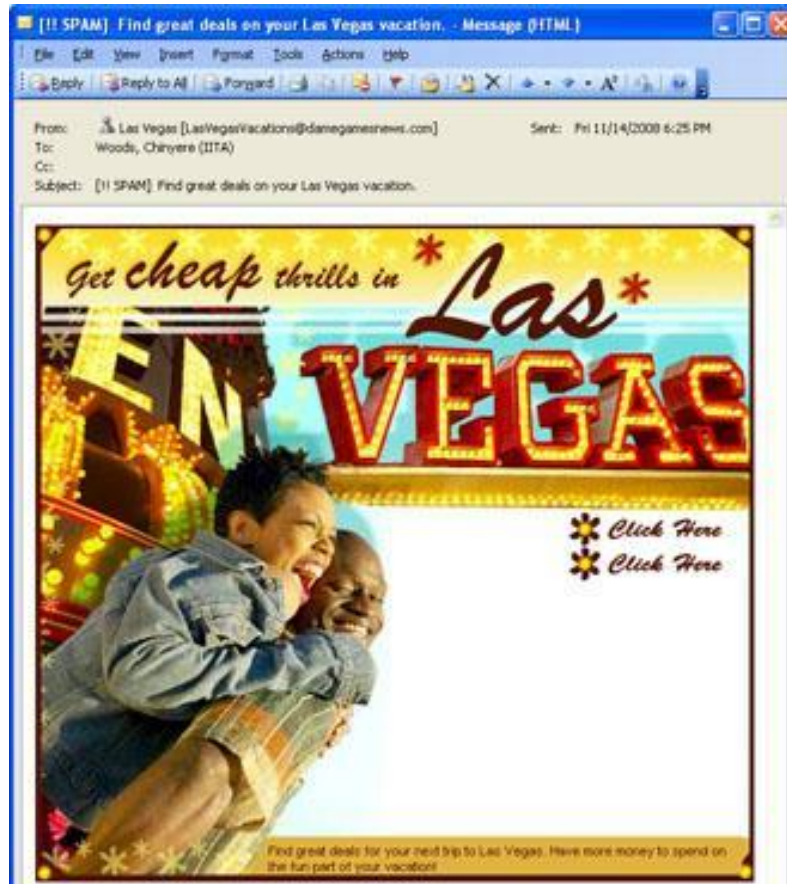


Figure 4: Example of image-based spam

Since these images make the messages much bulkier than simple text messages, the spammers end up using more bandwidth to send out fewer spam emails. With this new method, spammers also know that, to text filters, scanning for spam words like Viagra or stock in the e-mail messages will be difficult because the words are imbedded in an image. The picture of a word like Viagra is different from the word itself. The human eye can recognize and understand these words imbedded in pictures while the computer or filter cannot. It is for this type of image-based spam that we will try to develop a filter to trap.

Text and image-based spam:

These are spam messages with a combination of the above two formats, that is, the message body contains both text and images. In some cases, the images that appear as one are actually a set of images arranged side by side to give the impression that it is just one image. For people that have 'automatically turn off pictures' in their mailboxes this is easy to spot. Figure 5a show what the spam image looks like before downloading the pictures while Figure 5b shows the message itself which appears as an image.



Figure 5: The same spam message in 5a after downloading the pictures

3. TYPES OF SPAM FILTERS

All spam filters available today act in one of a few ways, or a combination of these ways to filter out the spam messages from users' mailboxes. The perfect anti-spam filter is yet to be invented and perhaps that explains why combinations of methods are used. Like earlier noted, whenever new measures are put in place to protect against spam, the spammers try to come up with new ways to get around the new measures. Two basic terms common in spam filtering are:

- **False Positive** - This is where a spam filter identifies a message as spam when it is innocent. This is the worst kind of error for a spam filter to make. It is better to err on the side of false negatives.
- **False Negative** - This is where a spam filter fails to identify a spam message as spam. A lesser problem than false positives, but still to be avoided.

The "ideal" spam filter would produce zero false positives and zero false negatives.

Content Based Filters

These types of filters analyze the contents, subject and headers of a message to try to determine if it is spam, usually by looking out for certain words. For this to be possible it means that the whole email must be received before any scanning/ filtering can be done. There are really three approaches to this: expert systems, also known as heuristics (clever people read huge amounts of spam to determine patterns that are updated relatively infrequently—e.g., SpamAssassin); fingerprint-based with frequent updates (spam messages are collected either from honey pots); and machine learning, usually Bayesian (the anti-spam filter "learns" by being shown examples of spam and non-spam). Most content-based anti-spam schemes try to identify spam by in some sense "understanding" the text contained in the email message. As a result, content-based filters are best deployed as "close" to the recipient as possible. This is especially important when using Bayesian filters. For example, for most people a mention of Viagra in a message is a good indicator of spam.

Over the years, spammers have been aware that their messages were being killed by these content filters and have resorted to ever more desperate tricks to try to fool the content filters. This would explain why you get so much mail for "Vi@gra", "Mort.gage", "L|O|a|n|\$." This practice has become so prevalent nowadays that older content-based filters have been performing less well. However, some of the more modern offerings have the ability to perform "wildcard" searches and at least one has the ability to "see through" the spammer's attempts at "obfuscating" the words/phrases such as in the examples above and even recognize that these attempts are being made - therefore indicating that it is spam.

Advantages:

They are flexible. You can easily tailor the filtering to the exact type of spam message you have to deal with and, just as importantly, not to balk at words or phrases that you use daily in your business or with your friends.

Disadvantages:

They require more "hands on" tuning and maintenance. As spammers resort to new tricks to foil the filters, or new products get advertised, extra filters have to be created to deal with them.

Bayesian Filters

This is one good example of a content-based filter. This method was developed in 1702 by Thomas Bayes. He developed a formula which allowed him to determine the probability of an event occurring based on the probabilities of two or more independent evidentiary events. Bayesian filters have to be "trained" from known "good" and "bad" e-mails, this will enable them to filter spam emails. During training they extract "tokens" (separate words) and store them in a database. When analyzing a new message, the message is split into tokens and each token is given a value according to the following criteria:

1. The frequency of the token in spam messages that the filter has been trained on
2. The frequency of the token in good messages that the filter has been trained on
3. The number of spam messages the filter has been trained on
4. The number of good messages the filter has been trained on

In basic terms, Bayes' Formula is based on a mathematically formula that is represented as:

$$P(E_j | F) = \frac{P(F | E_j)P(E_j)}{\sum P(F | E_i)P(E_i)}$$

- Assuming that we are making reference to three independent evidentiary events, we have the probability that the main event will occur to be:

$$\frac{abc}{abc + (1-a)(1-b)(1-c)}$$

- The formula can be expanded to accommodate any number of evidentiary events.

In practice, the first thing a Bayesian filter does is split the email message into tokens and build a table of all the tokens it intends to use in the decision-making process. Tokens generally refer to words or a group of characters separated by a space or special characters. Once the Bayesian filter has the list of tokens in the message, it searches the spam and non-spam token databases for these tokens, calculates each token spamicity. Depending on what the resulting figure, the Baye's filter can then determine if the email message containing of the tokens/ words are spam messages. However, they can be used to filter only text-based email messages. Because it is measuring probabilities, the Bayesian approach considers all the evidence in the email, both good and bad. Words that occur disproportionately rarely in spam (like "though" or "tonight" or "apparently") contribute as much to decreasing the probability as bad words like "unsubscribe" and "opt-in" due to increasing it. So, an otherwise innocent email that happens to include the word "sex" is not going to get tagged as spam.

Advantages

One great advantage of the statistical approach is that a person does not have to read so many spam messages. To recognize individual spam features is very difficult, because you have to try to think like the spammer. But the real advantage of the Bayesian approach is that you know what you are measuring. They also require less maintenance than other filters. Once the engine has been "trained", they pretty much look after themselves. They automatically adapt to shifting trends in spam. Because Bayesian filters continue to learn from newly arrived messages, they will naturally adapt to shifting trends.

They will automatically adapt to the particular user's usual e-mails. If a user is, for instance, a loan officer, then messages that repeatedly mention loans will not necessarily be identified as spam. They also have good record of minimum false positives.

Disadvantages

Bayes filtering is only as good as the messages on which they are "trained". Many filters based on this technology come "pre-trained", but obviously not on your type of messages. All will require some time before they reach optimum filtering ability. They have the potential to be fooled by diluting the spam message with enough obviously innocent words as possible.

Address Based Filters

These types of filters scrutinize the address of the sender of the email to determine if the message is a spam or not. They usually will not analyze the content of the message and that is why nowadays are rarely used on their own, but are still used as part of an integrated filtering system comprising some of the other methods like any content-based filter for effective spam filtering. There are mainly three different approaches to address based filtering.

Blacklisting

Domain Name Server 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. Blacklist filters, conversely, will allow messages from any address unless the address is on a list of known "bad" sources.

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 other side however, 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. This sometimes happens when the people that compile the real-time blacklists available on the internet put entire ranges of IP addresses on their blacklist even though previous abuse occurred only in a certain part of that range. This results in "collateral damage" - the situation where innocent people get blocked as a by-product of stopping the spammer and is the subject of much contention.

Whitelisting

Whitelists are the exact opposite of blacklists. Whitelist filters will not accept e-mails from any address unless it is on the list of known "good" e-mail addresses. Instead of filtering possible spammers, they keep track of secure IP addresses or email addresses which are lists of contacts that the mailbox user deems are acceptable to receive emails from without scrutiny. An email whose sender is on such a list will be accepted with no further filtering. One good advantage of whitelists filters is that they are guaranteed to stop e-mail from unwanted sources. However, if such lists are set incorrectly, they may reject legitimate emails.

Challenge/Response Filters

Challenge/Response filters are characterized by their ability to automatically send a response to an unknown sender asking them to take some further action to ensure their message will be received. In this scheme, no mail gets through unless it is whitelisted by the recipient. If a sender tries to deliver to a protected mailbox, the message in question is held in a quarantine queue and a challenge is returned to the sender. This can be as simple as "reply to this message," to sending them an image or a code and asking them to enter the word that is included there. Once the challenge is correctly solved, the sender address is added to the recipient's whitelist and the original message is delivered.

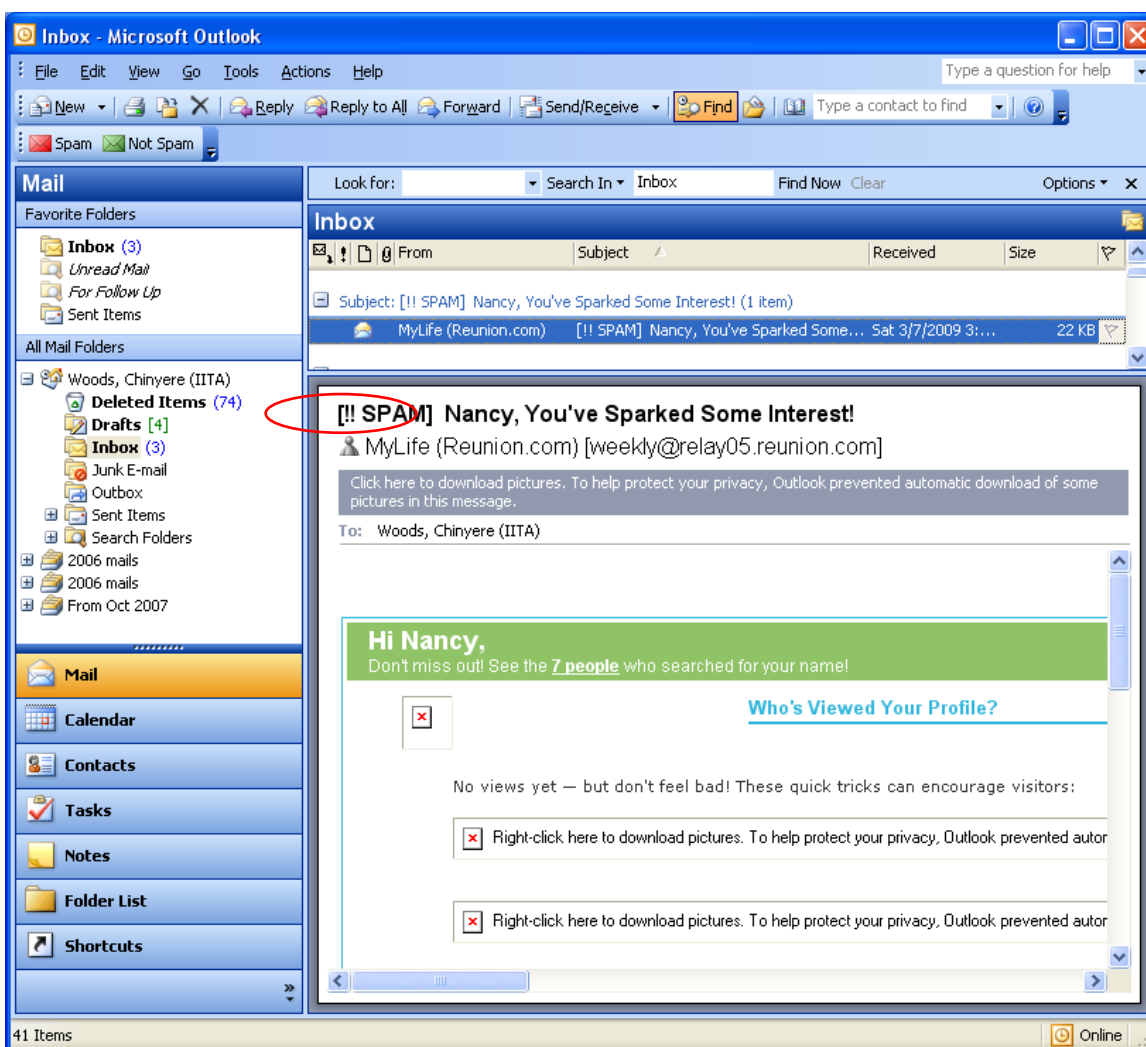


Figure 6: A request by the anti-spam filter for the mailbox user.

Although these tend to be very effective, there is anecdotal evidence that some users are confused by the confirmation requests; this is particularly problematic for non-technical users. Eudora is an Antispam that uses this method. It allows in only messages from people on the users address book. If anyone else writes, Eudora auto-replies with an apology and tells the writer to try again with a certain code word in the subject line. Once Eudora sees the code word in the subject line, it accepts the message without fuss. Since spammers are not patience enough to reply, they will not re-send the message anyway.

Community Filters

These types of filters work on the principal of "communal knowledge" of spam and communicate with a central server. When a user receives a message that is spam, they simply mark it as such. This information is posted to the central server where a "fingerprint" of the message is added to the database. When enough people have "voted" the message as spam, it will be blocked from user's inboxes in the future. Below is an example of a mailbox asking the user to classify a message as spam or not. The red mark shows the systems request.

Advantages

These types of filters are easy to set up and require very minimal administration. By marking a message as spam, that message is prevented from being delivered to thousands of others.

Disadvantages

Before enough votes are cast for a particular message, people will be getting the spam messages. Also, one person's idea of spam may not be another's - consequently, some innocent mails may be blocked by over-zealous people preventing their delivery. Some spammers will slightly change each message sent so that the "fingerprint" of the message is different, meaning it may not be recognized as the previous message blocked.

4. SPAM FILTERING TECHNIQUES/ FILTER PLACEMENT

Spam filtering has been in existence since the infiltration of spam messages in the Internet. There are different approaches and methods available for use. One approach used by SpamAng, as proposed by Longe O.B. is for mail servers to scan and filter outgoing messages before they are sent off. This will greatly protect that particular mail server/ DNS from being blacklisted as well as reduce the number of spams on the Internet.

However, the common approach used by most anti-spam filters for now is to tackle the spam messages from the receiver's end. With this approach, the mail server scans and filters possible spam before sending the ham messages to the individual mailboxes. In some cases, the server will make a list of the filtered messages and send this list to the user to confirm and eventually delete the list (figure 3). In some cases, the spam email actually arrives at the receiver's mailbox but are flagged and moved to a separate folder tagged spam emails. However, some important emails also get sorted into this folder. One final distinguishing factor to consider between spam filters is their actual placement. There are three main schemes for filter placement:

Filters that Integrate with your e-mail client

Many modern spam filters will integrate with popular e-mail clients, such as Outlook or Outlook Express to try and filter the messages in the mailbox. The advantage of this method is that there is minimal impact on the users' normal e-mail reading habits. Spam messages are usually simply moved to a "Junk Mail" folder where they can be reviewed and/or deleted. On the other hand, this method ties one to one current e-mail client. They are also inflexible, for instance, when running Microsoft Outlook with an integrated spam filter, whenever a spam message arrives, you still get an alert that a new message has arrived. You have to go into the Outlook interface to confirm that the newly arrived message was spam and not an important e-mail. This will result in either ignoring a new message that arrives, or continually being disturbed only to find that the new message is spam.

Filters as a "proxy" between the mail server and an e-mail client

These filters run in the background on the desktop and periodically poll one's e-mail server, retrieve the messages found and act on them before they reach your normal e-mail client. Advantages include: - Flexibility - usually have more control over the messages on your server and can mark, move or even delete messages before they are seen by your normal e-mail client. Security - they represent another layer between the internet and your e-mail client. They usually will not run any applications or run scripts found in the e-mail message. Disadvantages are that they impose a change on one's normal e-mail viewing habits like effective use of these involve turning off auto-checking on your normal e-mail client so that the proxy has a chance to work on the server first.

Server Based Filters

These are usually only used in a corporate or business environment rather than in the home. All e-mail arrives at a central email server where it is filtered by the server-based filter and individual users collect their messages on their desktop from the central server. Advantages include a central management of all e-mail filtering rules ensuring consistency across the network. Individual users have no little or no responsibilities for spam management, freeing them to be more productive in their work. The disadvantages are usually that they require more maintenance and require the presence and time of an experienced network administrator to manage the filter. They are often more expensive.

Efforts on fighting Spam

Spam filters have now become an integral part of Internet usage. Offering easy and user-friendly methods of keeping unwanted mail away from the computer, these filters have become immensely popular amongst the users of the World Wide Web. Various people/ companies have employed various ways to fight spam. The most common way to tackle spam is at the receiving end. That is to say that most anti-spam filters work on inbound email messages to filter them. Recently though, some companies now go for outbound spam filtering, usually to protect the company's image because if it is known that spammers operate from the company's domain, it's network address could be blacklisted.

Throttling

The throttling method is an interesting way to fight spam attacks. The throttling mechanism is sensitive to the extraordinary traffic activities originating from a single network or host. Spammers send e-mails in big quantities, and throttling mechanism slows down this spamming activity, since a certain amount of bandwidth is allocated to a single network.

There are cases that a legitimate mailing list may send out huge quantities of mail, but each message is addressed to different users on different networks. Throttling also prevent spammers using dictionary attack to find valid e-mail addresses on the network.

Content Filters

While origin-based filters such as whitelists and blacklists examine e-mail headers and other network information, content filters detect spam by looking inside the e-mail and examining the message contents. Most content-based spam detection systems try to “understand” the text to various extents in order to identify spam (Allman 2003). A simple word filter, for example, could look to see if the message contains the words Viagra or sex or the phrases “buy now” or “you’ve won” to determine whether it is spam. Filters based on this technique are commonly called keyword-based filters.

Bayesian Filters

Spam filtering systems using Naïve Bayesian classification were originally proposed by two separate parties at the AAAI-98 Workshop on Learning for Text Categorization; one was by Pantel and Lin (1998) and the other was by Microsoft Research (Sahami et al. 1998). Bayesian (also known as statistical) filters work by analyzing the words inside an e-mail message to calculate the probability that it is spam. This probability is based on not only those words that provide evidence that a message is spam, but also on those words that provide evidence that a message is not spam. Words that are not generally found in spam messages contribute to the probability value in very much the same way as words that are frequently found in spam messages (Graham 2002).

Rule-Based Filters

Rule-based, or heuristic, filters search the e-mail message for patterns that indicate spam. These patterns could include specific words or phrases, malformed message headers and large amounts of exclamation marks and capital letters. Detection of a specific pattern attributes an amount of points to an e-mail message and once the point value of an e-mail exceeds a set threshold, it is classified as spam. Rule-based filters were the most common type of spam filter until 2002, when Bayesian filters became popular. The challenge with rule-based filters is that, since the rule set is largely static, they are easily defeated by spammer techniques such as word obfuscation. For example, a rule-based filter that checks for the word Free will not be able to detect the string F*r*e*e as spam. It is incredibly difficult to include rules for every possible misspelling of common spam word, which has limited the effectiveness of such filters.

Other Filters

While the above filters are the most commonly used spam filtering techniques, it is by no means an exhaustive list. Other data mining, machine learning and text classification techniques currently under research include: digest-based filters (Damiani et al. 2004), density-based filters (Yoshida et al. 2004), Chi-squared filters (O'Brien & Vogel 2003; Zdziarski, 2005; Zan & Serano, 2005), global collaboration filters (Hulten et al. 2005) and artificial neural networks (Drewes 2006). Social network techniques, such as Reputation Network Analysis are also under investigation (Golbeck & Hendler 2004). It should also be noted that the various spam filtering methods described above are by no means mutually exclusive. A popular spam filtering program, SpamAssassin (SpamAssassin, 2005), uses a combination of Bayesian filtering, rule-based filtering and blacklist checking to calculate a spam score for a particular e-mail message. Messages that exceed a particular user-defined threshold are then marked as spam and dealt with appropriately.

Intrusion Detection System

Rajkumar et al (2006) proposed a technique which revolves around using intrusion detection system (IDS) techniques, specifically audit log analysis, to stop spam messages before they enter the network. Network precursors such as port scans are used as evidence that a particular IP address will soon send spam. Once enough evidence against a particular IP address is gathered, rules will be added to the network firewall to block SMTP connections from that address. They term this system a domain specific dynamic blacklist (DSDBL), as it essentially uses network information to dynamically blacklist the IP addresses of spammers attacking ISP domains. This technique aims to stop spam at the network gateway.

Filtering by Number of Recipients.

Chih-Chien (2003) pointed out that filtering by number of recipients involves screening out e-mail being sent to a large number of recipients. This method is based on the assumption that junk e-mails generally are sent in bulk. However, this method is easily defeated by bulk e-mail programs that automatically reconnect to the mail sever or automatically change the e-mail title and/or body messages after sending a certain number of e-mails. The spammer can set this number to as little as twenty or less. Additionally, this method assumes all bulk e-mails to be junk, and thus may filter out important and solicited e-mail as well as unsolicited e-mail.

Filtering by keyword

Keyword filtering is another method of screening junk e-mail. Several e-mail service providers currently provide content-based filtering on their mail servers. However, this filter method also does not perform perfectly, and thus users still must check filtered e-mails and put any unsolicited e-mail in the junk e-mails may contain specific words. "Sell", "on sale", and "sex" are some typical keywords used to filter e-mails with this approach. These words should be defined in advance. However, this approach suffers the problem of high possibility of solicited e-mails being filtered out because they happen to include words on the filter list (Pazzani, 2000; Sakkis et al, 2003).

Filtering by sender address

Filtering Spam out based on the sender address is the most straightforward approach (Cranor and LaMacchia, 1998). However, the problem with this technique is that it is easy for senders to obtain new e-mail addresses or use fake e-mail addresses. Besides, defining Spammers and separating them from the normal users represent another problem. Disputes can occur if mail servers improperly classify normal users into Spammers and refuse to deliver mails sent by these normal users.

Filtering by address validity

The format of e-mail is described in the Request for Comments (RFC) 822 initially designed by Crocker (1982). According to RFC 822, each e-mail must include the headers of "From" and either "To", "CC", or "BCC". The sender address is listed in the "From" header and the receiver addresses are listed in the "To", "CC", or "BCC" headers. All e-mails should have at least a sender and a receiver. Although RFC 822 mandates the existence of the "From" header in e-mails, no mechanism exists to ensure the validity of the sender address. That is, the e-mail address left in the "From" header can be invalid if the sender so chooses. Chih-Chien (2003) said that the validity of e-mail addresses of the header session may provide a cue for anti- Spam filtering. The "To", "CC", and "BCC" headers are used to identify message recipients.

"To" is used to identify main messages rather than main message recipients. Moreover, "CC" standing for Carbon Copy, identifies second message recipients rather than main message recipients. The function of the "To" and "CC" headers thus are very similar. However, "BCC" differs from the "To" and "CC" header. "BCC" means Blind Carbon Copy and blinds individual recipients to identify other recipients. No receivers thus can see the e-mail addresses of other "BCC" receivers. E-mail servers generally add "undisclosed recipients" to the "To" header if only "BCC" is unavailable. Junk e-mails are sent in bulk and have numerous recipients. Spammers generally use the "BCC" header rather than "To" or "CC" for the receiver address to avoid revealing that the mail part of a large batch. This habit of Spammers provides a further cue for judging the possibility that an inbound e-mail is junk.

Heuristics

Heuristic, or rule-based, analysis uses regular expression rules to detect phrases or characteristics that are common to spam; the quantity and seriousness of the spam features identified will suggest the appropriate classification for the message. The current popularity of this technology has largely been driven by its simplicity, speed and consistent accuracy. Furthermore, it is superior to many advanced filtering technologies in the sense that it does not require a training period. However, in light of new filtering technologies, it has several drawbacks. It is based on a static rule set and the system cannot adapt the filter to identify emerging spam characteristics. Therefore, if a spammer can craft a message to penetrate the filter of a particular vendor, their messages will pass unhindered to all mail servers using that particular filter. Open-source heuristic filters provide both the filter and the rule set for download, allowing the spammer to test their message for its penetration ability.

5. DIGITAL IMAGES

Pictures are the most common and convenient means of conveying or transmitting information. A picture is worth a thousand words. Pictures concisely convey information about positions, sizes and inter-relationships between objects. They portray spatial information that we can recognize as objects. Human beings are good at deriving information from such images, because of our innate visual and mental abilities.

About 75% of the information received by human is in pictorial form (Minakshi, 2017). A digital remotely sensed image is typically composed of picture elements (pixels) located at the intersection of each row i and column j in each K bands of imagery. Associated with each pixel is a number known as Digital Number (DN) or Brightness Value (BV), which depicts the average radiance of a relatively small area within a scene (Figure 6.1). A smaller number indicates low average radiance from the area and the high number is an indicator of high radiant properties of the area.

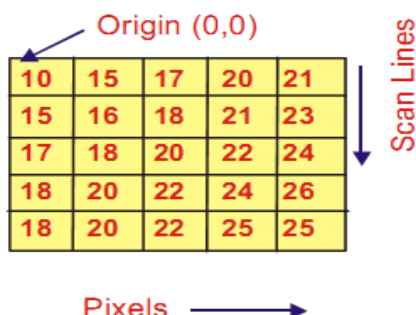


Figure 7: Structure of a digital image

Concept of an Image

An image is a collection of numbers that constitute different light intensities in different areas of the image (Saurabh and Gaurav, 2010). This numeric representation forms a grid and the individual points are referred to as pixels. Most of the images consist of a rectangular map of the image pixel known as the bits where each pixel is located and its color. These pixels are displayed horizontally row by row. The number of bits in a color scheme, called the bit depth, refers to the number of bits used for each pixel. The smallest bit depth in current color schemes is 8, meaning that there are 8 bits used to describe the color of each pixel (Saurabh and Gaurav, 2010). Monochrome and grayscale images use 8 bits for each pixel and are able to display 256 different colors or shades of gray.

Digital color images are typically stored in 24bit files and use the RGB color model, also known as true color. All color variations for the pixels of a 24-bit image are derived from three primary colors: red, green and blue, and each primary color is represented by 8 bits. Thus in one given pixel, there can be 256 different quantities of red, green and blue, adding up to more than 16 million combinations, resulting in more than 16 million colors. Not surprisingly the larger amount of colors that can be displayed, the larger the file size (Saurabh & Gaurav)

Image Processing

The majority of today's steganographic systems uses images as cover media because people often transmit digital pictures over email and other Internet communication (for example, eBay). Moreover, after digitalization, images contain the so-called quantization noise which provides space to embed data (Westfeld and Pfitzmann, 2021). In digital image processing systems, one usually deals with arrays of numbers obtained by spatially sampling points of a physical image. After processing, another array of numbers is produced, and these numbers are then used to reconstruct a continuous image for viewing (William, 2021).

Image Compression

Image compression is a commonly used method for reducing the file size of an image, without reducing the aesthetic qualities enough to become noticeable by the naked eye. Broadly speaking, it extracts all the information from an image that the human eye is not perceptible to and would therefore not miss should it not be there. The compression of JPEG images contains several processes:

- i. Colour space transformation (Converting pixel values to YCbCr)
- ii. Downsampling the chrominance values
- iii. Transforming values to frequencies
- iv. Quantization
- v. Zig-Zag ordering
- vi. Lossless Compression

a. Colour Space Transformation

The first step is to convert the RGB colour layers of the image into three different components (Y, Cb, and Cr). The Y component relates to the luminance (brightness), and the U and V elements relate to chrominance (colour). The chrominance coefficients of an image (Cb and Cr) are determined by a 2D grid that has blue to yellow on one axis, and red to green on another.

b. Downsampling the Chrominance Values

This phase is to downsample (or subsampling) the image. It is a common belief that the human eye is more sensitive to changes in brightness than to changes in colour. This means that it is possible to remove a lot of colour information from an image without losing a great deal of quality. As a result of this, much of the compression takes place by downsampling the chrominance data to reduce the overall file size. JPEG compression often downsamples by taking four adjacent pixels and averaging them to one value. Doing this for the entire image will remove roughly half. All of this can be done without any noticeable drop in the quality of the image.

c. Transforming Values to Frequencies

The Discrete Cosine Transform (DCT) is used for JPEG images to transform them into frequencies. DCT is a mathematical transform (typically a cosine function) that converts the pixels by seemingly 'spreading' the location of the pixel values over part of the image.

It does this by grouping the pixels into 8 x 8 blocks and transforming them from 64 values into 64 frequencies (DCT coefficients). By modifying just a single DCT coefficient, the entire 64 pixels in that block will be affected.

d. Quantization

The next step is arguably the most important one when compressing JPEG images. The aim is to quantize the values that represent the image after previous stage. Quantization is the process of taking the remaining coefficients and dividing them individually against a pre-determined set of values and then rounding the results to the nearest real number value. The goal is to eliminate the high frequency (lower-right) values.

e. Zig-Zag Ordering

Most of the quantized DCT coefficients are fairly typical. There are only a few values that hold numbers other than zero - the majority will always be zeros. It is also typical to see the non-zero numbers in the upper left, and zeros as you get towards the lower-right corner. As this is the case, this stage of JPEG compression reorders the values using a 'zig-zag' type motion so that similar frequencies are grouped together.

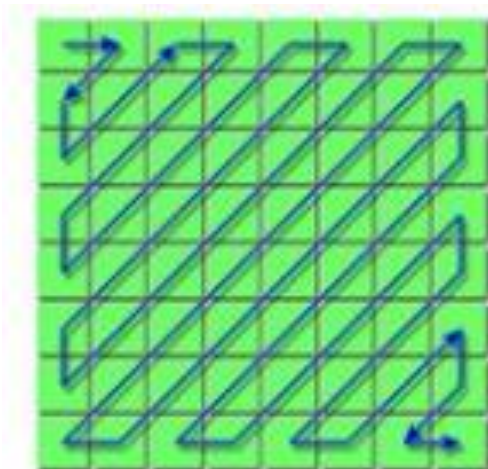


Figure 8: Zig-zag ordering of JPEG image components

2.7 Image-Based Filtering System Platforms.

Existing image spam filtering systems are mostly server-side filter systems maintained by trained administrators. Client-side filtering systems are designed to work without professional administration on top of an e-mail client. They can also be installed as a proxy between the e-mail client and the e-mail server. Most of these filters are based on a single or hybrid of the following approaches.

Table 1: Gaps Identified in Existing Image Spam Filters

S/No	Name of Image Spam Filter	Strength	Weaknesses
1	Visual Spam Filter	It is able to immediately detect if an image is steganographic and embedded with spam	Unable to detect phishing manipulation when the images are slightly manipulated by adjusting the compression level, colorimetry.
2	Image Content Analyzer	It is a signature-based technique that scales on familiar images with image-based filters that are traditionally	Using geometry and obfuscation confuses this filter and hence spammers can bypass email filters
3	Image Spam Detector		Still relies on blacklists and rules engine leaving a big gap for users to identify the spam,
4	Picture Spam Blocker	Works by searching for patterns in the headers or bodies of messages and the mail containing the image.	The underlying algorithm is not able to deal with stem words and its roots
5	Image Analysis Spam Filter	Its able to deal with block image Spam.	Not able to handle broken images i.e. splitting a single image

6. CONCLUDING REMARKS

Existing Image-based spam filters such as Visual Spam Filter, Image Content Analyzer Image Spam Detector, Picture Spam Blocker and "Image Analysis Spam Filter, utilize different techniques to detect and block spam that are conveyed through images, such as emails containing promotional graphics with embedded text. Such techniques include filtering based on suspicious image patterns, analyzing text embedded in images, identifying common spam image templates, and detecting unusual image sizes or file types associated with spam campaigns; essentially, these filters try to identify images that contain spam content even if the text within them are not readily visible to the naked eye. The challenge with these techniques is the difficulty in accurately detecting spam content embedded within images as the filters struggle to interpret visual information, leading to a higher rate of missed spam messages. This problem is exacerbated by the strength of the algorithms and models employed in designing the spam filters.

BIBLIOGRAPHY

1. Anjali A & Umesh L (2021), *A Secure Skin Tone Based Steganography Using Wavelet Transform*: International Journal of Computer Theory and Engineering, Vol. 10, No.1, February, 2021 1793-8201
2. Anon G (2018), *Audio Steganography: Overview*:
<http://www.snotmonkey.com/work/school/405/overview.html>
3. Artz D (2018), *Digital Steganography - Hiding Data within Data*:
4. http://www.cc.gatech.edu/classes/AY2003/cs6262_fall/digital_steganography.pdf
5. Balfrishan T and Amar P (2022), *Moderate Bit Insertion for Hiding Crypto-Data in Digital Image for Steganography*; A special issue from IJCA – www.ijcsonline.org
6. Bret D (2022), *A detailed look at Steganographic Techniques and their use in an Open System Environment*: paper from the SANS Institute Reading Room site. SANS Institute, 2022.
7. Brigitte S (2024), *Introduction to steganography*: Athabasca University, COMP607 Project, July 2024
8. Calpe A. (2023), *Steganography in Images*:
http://www.cs.ucf.edu/courses/cot4810/fall04/presentations/Steganography_in_Images.ppt
9. Christian C (2005), *Digital Steganography*: IBM Research Zurich Research Laboratory CH-8803.
10. Christian N, Bill E, Jay G, Morgan S, and Karli W (2018), *Wrox Professional C#*: ISBN 978-0-470-19137-8.
11. Cole E (2023), *Hiding in Plain Sight*: ISBN 416-236-4433 Wiley, John & Sons,
12. Debnath B, Poulami D, Samir K and Tai-hoon K (2019), *Text Steganography – A Novel Approach*: International Journal of Advanced Science and Technology, Vol. 3,
13. Debra A and Daniel S (2019), *Data Compression*: www.ics.uci.edu/%7Edan/pubs/DC-Sec1.html
14. Donal O, Micheal P and Hitesh T (2021), *Electronic Payment Systems for E-Commerce*: ISBN 1-58053-463-5, Artech House Inc.
15. Douglas C (2020), *Internet*: Microsoft Encarta 2020, Microsoft Cooperation

16. D. Ebem, Joseph Chinonye Onyeagba, G. Ugwuonah (2021): Internet Banking: Identity Theft and Solutions - The Nigerian Perspective. Law, Business, Computer Science. The Journal of Internet Banking and Commerce. <https://www.semanticscholar.org/paper/Internet-Banking%3A-Identity-Theft-and-Solutions-The-Ebem-Onyeagba/54822b71a2044498d4c3b448e905497c5fcf3b8b>
17. Longe, O.B, Asani, E. & Ashante D. (2021): A Maximum Entropy Classification Scheme for Phishing detection using Parsimonious Features. TELKOMNIKA (Telecommunication Computing Electronics and Control), SCOPUS Indexed/ScimagoJR indexed journal, Q2 on Electrical and Electronics Engineering, SJR: 0.283, CiteScore: 1.09, SNIP: 0.730.
18. Fabien P (2021), *The information hiding homepage Digital watermarking & steganography*: <http://www.petitcolas.net/fabien/steganography/>
19. Farid, H. (2020), *Detecting Steganographic Messages in Digital Images*: <http://www.cs.dartmouth.edu/~farid/publications/tr01.pdf>
20. Ingemar J, Mathew L, Jeffrey A, Jessica F and Ton K (2018), *Digital Watermarking and Steganography*: ISBN 978-0-12-372585-1 Morgan Kaufmann Publishers
21. Joel M (2020), *Murach's C# 2010*: ISBN 978-1-890774-59-2 Mike Murach & Associates, Inc.
22. Khosravi S, Abbasi D, Yektaie M and Khosravi M (2020), *A New Method to Steganography Whit Processing Picture in Three Colors (RGB)*:Khosravi, Sara,AbbasiDezfouli, Mashallah, Int. J. Comp. Tech. Appl., Vol 2 (2), 274-279
23. Mangarae A (2024), *Steganography FAQ*: http://www.infosecwriters.com/text_resources/pdf/Steganography_AMangarae.pdf
24. Minakshi K (2017), *Digital Image Processing*: Photogrammetry and Remote Sensing Division, Indian Institute of Remote Sensing, Dehra Dun.
25. Morkel T (2020), *An overview of Image Steganography*:Information and Computer Security Architecture (ICSA) Research Group.
26. Muhalim M, Subariah I, Mazleena S and Mohd R (2023), *Information Hiding Using Steganography*: Department of Computer System & Communication, Faculty of Computer Science and Information system, University Teknologi Malaysia, 2003.
27. Naeem Ahmed, Rashid Amin, Hamza Aldabbas, Deepika Koundal, Bader Alouffi & Tariq Shah (2022):_Machine Learning Techniques For Spam Detection In Email And Iot Platforms: Analysis And Research Challenges. <https://doi.org/10.1155/2022/1862888>, <https://onlinelibrary.wiley.com/doi/10.1155/2022/1862888>
28. Natarajan M and Lopamudra N (2022), *Steganalysis Algorithms for Detecting the Hidden Information in Image, Audio and Video Cover Media*: International Journal of Network Security & Its Application (IJNSA), Vol.12, No.1, 2022.
29. Katzenbeisser A and Petitcolas F (2021), *The Information Hiding Techniques for Steganography and Digital Watermarking*:Artech House Publisher.
30. Queirolo F (2018), *Steganography in Images*: <http://www.cse.buffalo.edu/~peter/cse741/Presentations/Refs/Queirolo.pdf>
31. Rupinder K, Mandeep K, Rahul M (2021), *A New Efficient Approach towards Steganography*: ISSN 0975-9646, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 2 (2), 2011, 673-676
32. Sahami, M.; Dumais, S.; Heckerman, D. & Horvitz, E. (2023): A Bayesian approach to filtering junk e-mail. AAAI Technical Report WS-98-05, Madison, Wisconsin, 2023.

33. Saurabh S. and Gaurav A (2023), *Use of image to secure text message with the help of LSB replacement*: ISSN 0976-4259, Invertis Institute of Engineering and Technology, Bareilly India.
34. Sellars D. (2018), *An Introduction to Steganography*: <http://www.totse.com/en/privacy/encryption/163947.html>
35. Shiva-Kumar K, Raja K, and Sabyasachi P (2022), *Hybrid Domain in LSB Steganography*: International Journal of Computer Applications Volume 19– No.7.
36. Tiwari R and Sahoo G (2023), *Designing Some Imperceptible Data Hiding Methodologies Using Steganographic Techniques*: International Journal of Information Technology and Knowledge management, Volume 1, No2, pp. 209-217.
37. William P (2021), *Digital Image processing*: ISBN 0-471-22132-5, John Wiley & Sons Inc.
38. Westfeld A & Pitzmann A (2021), *Attacks on Steganography Systems*: Department of Computer Science, Dresden University of Technology
39. Cryptography (2024) - <http://en.wikibooks.org/wiki/Cryptography>
40. Cryptography (2023) - <http://en.wikibooks.org/wiki/Cryptography/Introduction>
41. Data Compression (2019) - www.data-compression.com/looseless.html.
42. Internet - www.wikipedia.com/internetworld, 2024
43. JPEG (2021) - <http://en.wikipedia.org/wiki/JPEG>
44. IBM X-Force® 2020, Mid-Year Trend and Risk Report (August 2020).
45. IBM X-Force® 2021, Mid-Year Trend and Risk Report (September 2021).
46. Battista Biggio, Giorgio Fumera, Ignazio Pillai, Fabio Roli, Biggio, Battista; Fumera, Giorgio; Pillai, Ignazio; Roli, Fabio (2018). "A survey and experimental evaluation of image spam filtering techniques, *Pattern Recognition Letters*". *Pattern Recognition Letters*. 32 (10): Volume 32, Issue 10, 15 July 2018,
47. Li, Siyuan; Li, Ruiguang; Xu, Yuan; Zhou, Hao; Yan, Hanbing; Xu, Bin; Zhang, Honggang (2018-09-01). "WAF-Based Chinese Character Recognition for Spam Image Filtering". *Chinese Journal of Electronics*. 27 (5): 1050–1055. doi:10.1049/cje.2018.06.014. ISSN 1022-4653.
48. Aradhya, H., Myers, G., Herson, J. A. (2020). Image analysis for efficient categorization of image-based spam e-mail. In: Proc. Int. Conf. on Document Analysis and Recognition, pp. 914–918.
49. Wu, C.-T., Cheng, K.-T., Zhu, Q., Wu, Y.-L., 2022. Using visual features for anti-spam filtering. In: Proc. IEEE Int. Conf. on Image Processing, Vol. III, pp. 501–504.
50. Longe, O.B. (2011). On the use of Image-based Spam Mails as Carriers for Covert Data Transmission. *Computer & Information Systems Journal*. Vol. 15. Issue 1. <http://cis.uws.ac.uk/research/journal/vol15.htm> (DBLP Indexed)
51. Longe, O.B. & Chiemekwe S.C. (2008). Probability Modeling For Improving Spam Filtering Parameters. *Journal of Information Technology Impact*. Vol 8, No.1. (SCOPUS, SCIMAGO Indexed)
52. Battista Biggio, Giorgio Fumera, Ignazio Pillai, Fabio Roli, "Image Spam Filtering Using Visual Information", 19th Int. Conf. on Image Analysis and Processing (ICIAP 2023), Modena, Italy, IEEE Computer Society, pp. 105–110, 10/09/2023.
53. Fabio Roli, Battista Biggio, Giorgio Fumera, Ignazio Pillai, Riccardo Satta (2021), "Image Spam Filtering by Detection of Adversarial Obfuscated Text", Workshop on Neural Information Processing Systems (NIPS), Whistler, British Columbia, Canada,

54. Guilherme Ramos, Mirko Marras & Ludovico Boratto (2024): SIGIR '24: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. <https://sigir-2024.github.io/proceedings.html>
55. Fumera, G.; Pillai, I. & Roli, F. (2020): Spam filtering based on the analysis of text information embedded into images. *Journal of Machine Learning Research* (special issue on Machine Learning in Computer Security), 17:2699–2720, 2020.