Chapter - 6

Spam and Anti-Spam
Introduction

All Internet users are genuinely concerned about spam and other related online threats which are supported by empirical evidence. These can cause i) theft of personal data such as credit card information, ii) online fraud involving counterfeit websites, iii) loss of personal information through illicit access to computer systems, and iv) false or misleading representations in the market place [p-4]. Thus, the spam wastes time of users and Internet resources by way of consuming storage and bandwidth, besides costing money in combating it. E-Commerce cannot flourish in an uncertain environment where security of personal data is vulnerable. The aim of this chapter is to provide an introduction to various types of spam and analyze its objectionable features. Further, the problems caused by it along with the statistical information about the economic loss have also been discussed. This chapter also provides a detailed and categorized account of technological solutions which could be installed to control spam [p-10]. Moreover, it carries discussion about filter classification, a general filter model and various evaluation measures of filter performance.

6.1. Introduction to SPAM

The term “spam” reportedly came to be used in connection with online activities following a mid-1980s episode in which a participant in a MUSH created and used a macro that repeatedly typed the word “SPAM,” interfering with others’ ability to participate. It is also said that the nickname “Spam” apparently came from an annoying song in a Monty Python skit. In the skit, actors sing the word "spam" over and over again, becoming louder and louder throughout the skit, until none of the players can hear any of the others. Ultimately, the singing drowns everything else and the skit ends. The beginning of this phenomenon can be traced to mail sent as early as 1978 over the Arpanet network to all of its subscribers on the west coast of the United States. Spam [6.1] is unsolicited bulk commercial Communications which can be sent in a number of ways and with a variety of purposes. It has reached a point where it threatens the future development of e-commerce and the information society [6.2]. They degrade and depress the quality of virtual life. Spamming creates considerable impediments to the efficiency of e-mail.
systems, choking up Internet bandwidth and access rates, reducing efficiency and costing
Internet service providers and individual users lost time through their having to manage spam.

The increasingly sophisticated variants of spam and the threats [6.3] they pose have brought
anti-spam measures to the forefront of attention of several governmental agencies,
consumer groups and businesses worldwide. The law on spam is developing rapidly.
Spamming is the scourge of electronic-mail and newsgroups on the Internet. Spammers are, in
effect, taking resources away from users and service suppliers without compensation and
without authorization. Everyday there seems to be a new development in the area of spam law,
some in the legislatures and others in the courts [6.4].

Spam not only impinges on the privacy of individual Internet users but also creates
economic losses in terms of time and money. Spam affects a diverse range of stakeholders
ranging from customers, small and medium sized businesses to larger corporations and even
governmental agencies [6.5]. These emerging markets will be more heavily targeted with spam
delivered in local languages. It is predicted that growth of Asian character spam will increase by
up to 100 percent from current levels at 5 percent to around 10 percent [6.6].

6.2. Magnitude of the Problem

Spamming causes manifold problems. This wastes time to users opening spam with misleading
subject lines, wastes Internet resources trying to block it, consumes storage and bandwidth, and
costs money in fighting it [6.7]. Spam is mainly used to: i) advertise different types of goods,
services or ideas, ii) trick recipients into giving up their credentials, iii) deliver malicious
software, iv) cause a temporary crash of a mail server.

Spam covers wide range of topics including porn, financial and stock, health and drugs,
hardware and software, attracting a partner, getting a degree, religious, holidays, gifts, job
offers, threats, etc. The spam genres include those similar to legitimate mail like letters, memos,
order confirmations, offers, etc. Some spam e-mail contents are only links, linking to web pages
of spammers and some others contain graphics like pictures, or animations embedded or as an
attachment. The graphics often contain hidden links to spammer’s web sites.

Spam costs yearly $12 billion or around $1,910 per employee in lost time [6.8]. A survey
conducted in November 2004 estimates that the cost of deleting spam to American businesses is $22 billion per year. The average spam messages per day were 18.5 and the average time deleting them was 2.8 minutes. The costs of blocking spam in 2003 were $230 million and were estimated to cost $419 million in 2008 [6.9]. The gravity of the problem could be well realized by the fact that the FTC obtained a restraining order and a freeze on assets under the Can-spam Act in federal court against an individual and various companies alleged to have sold bogus "human growth hormone" products over the Internet through spam [6.10]. In June 2005, the FTC and the defendants settled the claims, with the defendant required to pay $485,000 in consumer redress to the FTC. If it is found that the financial statements provided by the defendant were inaccurate, another $5.9 million will immediately become due [6.11]. According to MessageLabs Intelligence: 2008 Annual Security Report [6.12] average Spam in years 2007 and 2008 remained respectively 84.6% and 81.2%. According to another report [6.13] Commtouch, a Sunnyvale, California,-based e-mail and Web defence group, reported that the volume of spam messages hit 94% of all e-mail sent in 2008. The lowest percentage of 2008 came in the fourth quarter, at 59%. Ferris Research, a San Francisco-based IT market research firm, estimates that spam will cost a total of $130 billion worldwide in 2009, of which $42 billion of that is incurred in the United States alone. A recent research [6.14] reported that existence of spam e-mails decreases not only the level of the Gross Domestic Product (GDP) by about 0.1% of Japanese economy which is about 464 billion Yen but also reduces labour productivity. Current trends indicate that after a decade spam will be more than 95% of the Internet transmission and Internet will then transmit vast amount of information but minute amount of solicited one's [6.15].

6.3. Spam: A Conceptual Analysis

Spam as a concept emerged when Internet infrastructure was used for carrying unsolicited commercial messages but could not be defined with precision. The difficulty in defining spam arises because of its varying purpose, scope and magnitude. Internet users and providers differ widely in how they define spam and other forms of objectionable e-mail [6.16]. Some people consider all advertisements or even all unwanted messages to be spam, while others try to define it in terms of existing acceptable use, policies, purpose and/or number. Spam is the
The fundamental objection to spam is rooted to its three chief characteristic features which are (i) unsolicited, (ii) bulk and (iii) commercial. There is near unanimity [6.17] amongst scholars that spam is unsolicited bulk e-mail but whether it is to promote commercial interest or not is always open to question. Some jurisdictions omit the word "unsolicited" altogether. Washington and Oklahoma's restrictions on falsified routing information apply to solicited as well as unsolicited e-mail messages. The term "unsolicited" is crucial in determining the liability of the sender [6.18]. A communication sent without any prior relationship with the recipient and for which recipient had not given any express or implied consent would be called unsolicited. It can also mean that the recipient had prior relationship with the sender but that has been terminated (opt-out). It is also said that an electronic message is unsolicited if the recipient did not request to receive the message or consent to the recipient of the message [6.19]. The definition of the term unsolicited will also depend on whether the legislation in which it has been defined has adopted opt-in or opt-out model [6.20]. On this basis alone, it will be decided whether recipient shall or shall not be treated as having requested to receive the message or consented to the receipt of the message merely because the electronic address of the recipient was given or published by or on behalf of the recipient.

The root of the objection to spam lies in its volume and not to its content. Spam is sometimes defined as messages sent in large quantities [6.21]. A single or separate but identical copies of message sent to a very large number of recipients would qualify as bulk. Messages with substantially similar content sent in a large quantity would be treated as bulk. The sender may send multiple copies either on his own or may use mail exploder [6.22] for making multiple numbers of copies. A "mail exploder" is a server that takes an incoming message and forwards copies of the message to multiple recipients.

The term "bulk" has been objectively defined in some legislations as sending of (a) more than 100 electronic messages containing the same or similar subject matter during a 24 hour period [6.23] or (b) more than 1000 electronic messages containing the same or similar subject matter.
during a 30 day period, or (c) more than 10,000 electronic messages containing the same or similar subject matter during a one year period [6.24]. This approach of arbitrarily prescribing a particular number has been criticized on the ground that by stipulating a certain threshold for a message being classified as bulk, the door is effectively opened to find ways to avoid crossing the threshold while still sending the same volume of messages. The term bulk may also be defined in terms of a message sending rate, such as sending messages to 1000 addresses during a twenty-four hour period. This approach would places a substantial burden on participating ISPs who would be required to maintain complicated logs of all received e-mails during a specified time period.

The objection underlying unsolicited e-mails is their commercial content. The term "Commercial" is generally defined in terms of message content rather than the sender's actual or presumed motivation for sending the message; a typical definition includes any message that promotes the sale of goods, immovable property or services. On the basis of their commercial nature, the bulk e-mails are classified as unsolicited commercial e-mails (UCE) and unsolicited bulk e-mail (UBC). Generally legislations in other jurisdictions have regulated UCE's [6.25].

6.4. Technological Perspectives of Spam

Spam is introduced at various places into the e-mail system by spammers for their illicit financial gains using a variety of techniques [6.26] and tools that include spoofing, spam botnets [6.27], open proxies, open mail relays, untraceable Internet connections, and bulk e-mail tools [6.28] for sending unsolicited bulk e-mail. Spammers are proactive and use dynamic spam structures that constantly keep on changing the spam structure to circumvent anti-spam procedures. The increase in the use of anti-spam procedures has led spammers to push more spam into the system in order to reach more and more users and guarantee themselves a huge profit. Spammers operate as a creative group who work through secret networks to meet and share e-mail addresses. The two most effective approaches to gather or harvest e-mails are to monitor Internet use and to ask for e-mail addresses. Some of the most popular methods of e-mail harvesting are guessing, purchased lists, legitimate e-mail lists, web pages, white/yellow page sites, web & paper forms, UseNet posts, web browser, hacking, user profiles, IRC and chat rooms. The significant increase in the spam volume mostly due to botnets has been the key
development in the year 2008 [6.29].

6.4.1. Variants of SPAM

Spam originally originated in Internet e-mail but various other services on the Internet that include Instant Messaging (IM), UseNet, Blogs, Web Pages, and Voice on Internet Protocol (VoIP) are also being affected by Spam. In recent times, spam has also spread in various wired and wireless telecommunication services (sometimes called wireless spam) like traditional telephone service, short messaging service (SMS) and multimedia messaging services (MMS) on mobile phones.

Spam in Instant Messaging (SPIM) [6.30] is the posting of unsolicited messages through instant messaging services. It has many characteristics similar to spam that include deceptive subject lines and false sender names. SPIM messages advertise and link to websites for pornography, services and products. SPIM can generate bots that simulate real IM users and often legitimate IM users fail to understand them and fall a pray to IM spam.

Spam Over Internet Telephony (SPIT) [6.31] is the transmission of unsolicited calls over Internet telephony or Voice over IP (VoIP) [6.32]. It is more serious than unsolicited calls in the traditional Public Switched Telephone Network (PSTN) which are limited in number because of relatively high cost of PSTN call. In Internet Telephony, call costs are much lower. The system can be easily programmed for spam applications as it is an IP based system and spammers can multiplex a number of calls on a single line.

Web Spam [6.33] or comment spam [6.34] also called blog spam [6.35] refers to the use of illegitimate means to boost the rank positions in search engines of some target pages. Content spam and Link spam [6.36] are the two main categories of Web Spam. Content spam adds irrelevant or remotely relevant words in target pages to fool search engines to rank the target pages high. Link spam is a spam of hyperlinks in a webpage that manipulate link based ranking algorithms of search engines like PageRank to increase the rank of a web site or page so that it is placed as close as possible to top of search results. Review spam [6.37] is a comment spam posted by spammers in an e-commerce website to promote or damage the reputation of some product or service. In this type of spam, spammers write undeserving positive reviews to
promote some object or negative reviews to damage its reputation.

6.5. Anti-Spam Procedures

Several anti-spam procedures have been proposed that try to tackle the problem of spam at various levels in the system [6.38]. These procedures propose the use of diverse technological, legal, social and economical solutions. Proposed anti-spam procedures can be broadly classified in two categories vis-a-vis technological and legislative as shown in figure 19. It is beyond the scope of this chapter to discuss various techniques involved in each of the stated categories; however, a summary of each technological anti-spam group is given in this section. A detailed discussion of Legislative anti-spam procedures is given in other chapter 10.

![Figure: Classification of Anti-Spam Procedures](image)

6.5.1. Protocol Based Anti-Spam Procedures

SMTP is simple in content and requirements; and minimizes information that must be included in the mail exchange and leaves functions such as authentication to other protocols and applications. This simple architecture makes SMTP easy to implement and use, but the spammers have misused these advantages and have exploited its underlying trust to target recipients with spam, hide their own identities, and conceal their tracks [6.39]. There is a very weak authentication/authorization mechanism in SMTP. E-mail messages are introduced into
Chapter - 6: Spam and Anti-Spam

the mail delivery system with no authentication indicating whether the e-mail was written by
the person who claimed to own the sender mailbox or not. There is no checking on the
consistency of the e-mail envelope and e-mail header. During each hop of transmission, mail
transfer agents could also make changes on the e-mail header/body. In order to overcome these
disadvantages of currently used e-mail protocol namely SMTP and ESMTP protocols, add on
protocols (See figure 20) that add security features to SMTP have been proposed by many
researchers. Some of these have been approved by IETF while other failed to get it. These add on
protocols aim at providing secure messaging that has ability to provide data integrity, data
confidentiality, data origin authentication and non-repudiation.

Proposed protocols that add security to existing mail protocols are shown in figure 20. These
include SenderID [6.40], Sender Policy Framework also called Sender Permitted Form (SPF)
[6.41], Sender ID Framework (SIDF) Certified Server Validation (CSV) [6.42], DomainKeys
Identified Mail (DKIM) [6.43], DKIM Sender Signing Practices [6.44], Bounce Address Tag
Validation (BATV) [6.45], Lightweight MTA Authentication Protocol (LMAP) [6.46], Reverse MX
(RMX) Designated Mailer Protocol (DMP), Digital Signature [6.47], Secure Multi-purpose
Internet mail Extensions (S/MIME) [6.48], Privacy-Enhanced Mail (PEM) [6.49], Pretty Good
Privacy [6.50] (PGP) [6.51] and sponsored Top Level Domain (sTLD) [6.52]. Add on protocols

Design and Development of Efficient Techniques for Securing E-mail System from threats
(M. Tariq Banday) Page 148 of 266
perform only domain authentication either using Digital Signature or using IP addresses and some like S/MIME besides providing domain validation also perform encryption. Add on protocols aim to let an e-mail system at the receiving end detect spoofing and are not strictly anti-spam procedures. However, since spamming involves spoofing these techniques can greatly help the receiving domain in classify mail.

Some anti-spam techniques that propose to amend the existing e-mail protocols by including addition step(s) in the mail sending process are payment based solutions or challenge-response systems also called interactive filters. These procedures include zmail protocol [6.53], Differential e-mail delivery [6.54], proof of work [6.55] and CAPTCHA [6.56] techniques. Protocol change based anti-spam procedures although provide near complete solutions to the problem of spam but have failed to gain vide acceptability as these procedures require a major change or replacement of the existing e-mail protocols. Further, the backwards-compatibility challenge and the need for widespread, if not universal, adoption of any such solution, impede the effort to revise SMTP to overcome the treats to the current e-mail system. Thus far, e-mail software vendors have not sought to fix the spam problem within SMTP; rather, their solutions treat the protocol as given and use various other anti-spam procedures.

Several groups of anti-spam procedures that do not require a complete change in the existing e-mail protocols but suggest a local change in the incoming or outgoing or transmission process of the e-mail message have been proposed. These procedures include Rate Throttling also called shaping or economic filtering, Behaviour Analysis and information hiding also called Identity Hoping or Address Obscuring Techniques (AOTs). Various techniques which are based on Rate Throttling procedures aim at blocking high volumes of e-mail messages send in a particular time slot by slowing down the message that are likely to be spam [6.57]. Teergrubing [6.58] is the process of delaying the receipt of a message. As the sending server connects to a Teergrubing receiving server to deliver a message, the later delays answer request. A similar approach [6.59] uses the spam score calculated by the client filter to proportionately delay the delivery of the e-mail message based on its spam likelihood. Spammers in contrast to legitimate e-mail senders are more likely not to respond to retry, once the mail is bounced back or delayed. This behaviour of spammers has led to another anti-spam procedure named Greylisting [6.60] in which unknown senders are initially denied connection requests. Rate throttling can effectively
limit spam as they establish a feedback loop, however calculating the likelihood probability of an e-mail being spam is dependent on the efficiency of the client filter. These approaches require change of the e-mail clients and also will cause an annoyance to the legitimate senders. These approaches also raise issues for sending bulk legitimate mail and stopping spammers from acquiring sufficient resources like use of botnets. Reputation systems [6.61] can be used to classify e-mail messages on the basis of past spammers sending behaviour of the sender. Twining [6.62] proposes an anti-spam procedure that uses past behaviour of sender for message classification. It proposes guaranteed delivery of legitimate mail by using high priority queue for e-mail send by senders with good reputation and low priority queue for others. A research work [6.63] proposes to use separate mail transfer agent for error message received when replying to spammer's mail to identify spammers. Incoming as well as outgoing e-mail messages can be controlled by e-mail service providers at local clients/servers by various other techniques [6.64] as well. Information hiding methods are also called identity hopping methods which are Address Obscuring Techniques (AOTs) aiming to either hide e-mail address [6.65] or restrict them to limited use. These procedures include use of one time e-mail addresses [6.66], Single purpose addresses [6.67] with encapsulated policy that defines the acceptable use of the address e.g. whether the sender is allowed to directly send e-mail message to this address or not. Tysoyoshi Abe [6.68] has proposed an ad-hoc e-mail address service system that employs a cryptographic algorithm for generation of ad-hoc addresses. The addresses are associated with mail filter rules and subscriber's original address remains undisclosed to public. This algorithm enables mail transfer agents to filter mails by extracting rules from ad-hoc addresses. Addresses generated in AOTs are difficult for humans to handle which limit its use. Besides this, disadvantage AOTs suffer from complicated first time person to person contacting as a recipient cannot anticipate all regular e-mail connections.

6.5.2. Filtering Based Anti-Spam Procedures

A widely adopted solution to the problem of spam is filtering which automatically classifies incoming e-mail message into spam or legitimate mail without human interaction unlike interactive filters used in challenge response systems is Spam Filtering. Filtering more precisely non-interactive filtering does not require any fundamental change in the existing e-mail system.
and as such are common and popular solutions. Filtering solutions apply spam filtering algorithms at different points in the transmission process. It can be applied at servers (Sending or Receiving) or Routers [6.69] or the destination client. Filtering at the destination client solves spam problem partially as it reduces costs associated with productivity loss for the human recipient but does not prevent resource misuse. Filtering solutions close to the sending client will be more efficient as besides filtering spam it also prevents resource misuse. Solutions at the servers or routers partially prevent resource misuse but are at an advantage of providing centralized maintenance and administration. Client based filtering solutions have an advantage of correcting misclassification and tune rule set. For Corporate these solutions can be operated either on-site or off-site and the solution may be either hardware based or software based. On-site solutions are hardware or software filters installed within the organization. In off-site solution Mail Exchange (MX) records of the subscribing organization are redirected to an off-site filtering vendor who redirects the filtered mail to the subscriber. On-site solutions provide greater control and customization options to local administrators besides avoiding any possible security threat of e-mail being misused by the vendor. Off-site solution is easy and quick to install; it is also cost effective as no filtering hardware or software is required. Hardware filters are hardware appliances dedicated for the purpose of spam filtering. On the other hand software filters are commercial and open source software products that operate at either server or client level. Hardware filters are easier to deploy than software counterparts and are secure against operating system vulnerabilities as they may have operating systems in their firmware.

6.5.3. Filter Function

An e-mail message consists of two parts, namely header and body. Header is a structured set of fields, each having a name and specific meaning. It includes fields namely From, To, Subject, CC, BCC, etc. Message body generally referred to as content of the message is usually text, possibly with HTML markup and MIME encoded attachments. Message analysis and filtering involves selection of features from header and/or body or from message as a whole relevant for analysis. A filter may check the presence of certain words or may consider the arrival of a dozen of substantially identical messages in a certain slot of time. In addition to this, a learning-based filter analyzes a collection of labeled training data which are pre-collected messages with
reliable judgment.

A spam filter in general is an application that implements a function to classify an incoming e-mail message as spam or legitimate mail using a particular classification method. Such a system implements the following function:

\[ f(m, \theta) = \begin{cases} 
C_{\text{spam}}, & \text{if classified as SPAM} \\
C_{\text{leg}}, & \text{otherwise} 
\end{cases} \]

Where, \( m \) is the message to be classified, \( \theta \) is a vector of parameters, and \( C_{\text{spam}} \) and \( C_{\text{leg}} \) are respectively spam and legitimate messages.

Most of Spam filters including the statistical spam filters use machine learning classification techniques wherein the vector of parameter \( \theta \) is the result of training the classifier on a pre-collected dataset which may be rebuilding itself with every new message. \( \theta \) for such filters can be defined as:

\[ \theta = \theta(X). \]

Learning-based spam filters treat the input data as an unstructured set of tokens, filtering can be applied either to the whole message or to any part of it. For this group of filters with two classes of messages: spam and legitimate mail, there exists a set of labeled training messages, each message being a vector of \( d \) binary features and each label being \( C_{\text{spam}} \) or \( C_{\text{leg}} \) depending on the class of the message. The training data set \( M \) once pre-processed in this way, can be described as:

\[ X = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \]

\[ x_t \in \mathbb{Z}_2^d, \quad y_t \in \{C_{\text{spam}}, C_{\text{leg}}\} \]

Where, \( d \) is the number of features used. \( x_t \in \mathbb{Z}_2^d \) is a new sample the classifier should provide a decision \( y \in \{C_{\text{spam}}, C_{\text{leg}}\} \), \( y_1, y_2, \ldots, y_n \) and labels and \( \theta \) being the training function.

6.5.4. Filter Classification

A recent research work [6.70] done by James Carpinter and Ray Hunt has classified filtering anti-spam procedures broadly into Machine Learning and Non-Machine Learning filters as depicted in figure 21.
Non-Machine Learning filters use regular expression style rules to identify predefined tokens or characteristics common to spam e-mails and accordingly classify it on the frequency and severity of the features identified. They may also use information about the sender for classification. Non-machine learning filters cannot be automatically trained and require regular updates to their rule set to adapt to new spam characteristics. These filters may be based on heuristics [6.71], blacklisting [6.72], hash [6.73], whitelisting, signature or traffic analysis [6.74]. Blacklisting/white listing is used as part of more complex anti-spam procedures.

Machine Learning filters can further be grouped as those which allow accurate classification independently called complete solutions and those which provide a partial solution and are intended to be used as a component of some other filtering systems. Adaptive filters [6.75] group e-mail corpus into several categories according to their content and then compare unclassified e-mail with each group to determine its likely class based on the calculated likeliness ratio. Trusted Network based filters create a user's trusted network of correspondents with an automated graph method [6.76] or from 'trust' scores assigned by users to people they know [6.77] for classification of unclassified mail. Different complete machine learning filtering approaches include unified model, previous likeness and ensemble. Bayesian filters [6.78], Sparse Binary Polynomial Hashing (SBPH) [6.79], Orthogonal Sparse Bigrams (OSB) [6.80],
Support Vector Machine (SVM) [6.81], Chi by Degrees Freedom [6.82], Artificial Neural Network (ANN) [6.83], Markov Random Fields [6.84] and Genetic Algorithm [6.85] based filters are the most prominent filters that fall under unified model category [6.86]. Memory based filters like k-Nearest Neighbour (kNN) [6.87], case based reasoning system and pattern discovery algorithms are machine learning filtering algorithms that compare incoming e-mail with previous examples and as such belong to previous likeliness group. In ensemble [6.88] group of machine learning filters collaborative approach and results obtained from multiple classifiers is used to evaluate an incoming e-mail message. A comprehensive survey of the learning-based anti-spam procedures with a rich bibliography can be found in the recent research work of Enrico Blanzieri and Anton Bryl [6.89]. Ensembles of classifiers can often perform better than any individual classifier; this performance advantage can be attributed to three key factors namely statistical, computational and representational. This group of filters combine candidate hypothesis generated by several machine learning algorithms and form an ensemble thus reducing the risk of selecting incorrect hypothesis. They compute hypothesis from several algorithms like Bayesian, ANN, etc. and calculate the resultant hypothesis with weights assigned to each.

6.5.5. Learning Filter Model

Regardless of the type, machine learning e-mail filter consist of learning and detection stages as shown in figure 22.

The learning stage uses a training set in the form of known spam and legitimate e-mails collectively called e-mail corpus. Features are extracted from each e-mail of corpus, which are then reduced by a feature reduction function. A training function calculates likelihood probability of each feature occurring in spam and legitimate e-mails. It also calculates the prior probability of each class i.e. spam and legitimate. The feature set along with their likelihood probabilities are stores in a library for use in the detection stage. A new e-mail message is parsed in the detection stage with respect to the features in the feature set and group probabilities of each group are calculated. If the total calculated probability of spam is greater than some predefined threshold value, the mail is classified as spam otherwise it is classified as legitimate. The feature set and feature probability library is updated with every new classified e-mail.
6.5.6. Evaluation Measures

To evaluate the predictive accuracy of classifiers several measures have been proposed in literature [6.90]. The most simple measure is filtering accuracy namely percentage of messages classified correctly. More informative measures are recall and precision. Weighted error rate and weighted accuracy are a measure to evaluate filter accuracy. TCR is the relative cost of using the filter (and so having some false positives and some false negatives) to using no filter at all (and so having all the spam misclassified, but all the legitimate mail classified correctly). F-measure is the weighted harmonic mean of precision and recall. Since false positive are often more expensive than false negative, it is vital to compare the false positive rate of the classifier. The Receiver Operator Characteristics (ROC) curve is a graph to plot false positive against true positive, in which various threshold values are compared. These measures in terms of threat and normal (spam and legitimate) e-mail are shown in table 16. Let $n_{N \rightarrow N}$ be the number of normal messages classified as normal, $n_{N \rightarrow T}$ be the number of normal messages misclassified as threat, $n_{T \rightarrow T}$ be the number of threat messages classified as threat, $n_{T \rightarrow N}$ be the number of threat messages misclassified as normal and $\lambda$ be the weight on the accuracy of the classifier.
### Table: Evaluation measures for Spam Filters

<table>
<thead>
<tr>
<th>Evaluation Measure</th>
<th>Evaluation Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>[ Acc = \frac{n_{N-N} + n_{T-T}}{n_{N-N} + n_{N-T} + n_{T-N} + n_{T-T}} ]</td>
</tr>
<tr>
<td>Weighted Accuracy</td>
<td>[ W_{Acc}(\lambda) = \frac{\lambda \cdot n_{N-N} + n_{T-T}}{\lambda \cdot (n_{N-N} + n_{N-T}) + n_{T-N} + n_{T-T}} ]</td>
</tr>
<tr>
<td>Error Rate</td>
<td>[ Err_{Rate} = \frac{n_{N-T} + n_{T-N}}{n_{N-N} + n_{N-T} + n_{T-N} + n_{T-T}} ]</td>
</tr>
<tr>
<td>Weighted Error Rate</td>
<td>[ W_{Err}(\lambda) = \frac{\lambda \cdot n_{N-T} + n_{T-N}}{\lambda \cdot (n_{N-N} + n_{N-T}) + n_{T-N} + n_{T-T}} ]</td>
</tr>
<tr>
<td>False Positive Rate</td>
<td>[ FP_{Rate} = \frac{n_{N-T}}{n_{N-N} + n_{N-T}} ]</td>
</tr>
<tr>
<td>False Negative Rate</td>
<td>[ FN_{Rate} = \frac{n_{T-N}}{n_{T-N} + n_{T-T}} ]</td>
</tr>
<tr>
<td>Recall</td>
<td>[ r = \frac{n_{T-T}}{n_{T-N} + n_{T-T}} ]</td>
</tr>
<tr>
<td>Precision</td>
<td>[ p = \frac{n_{T-T}}{n_{N-T} + n_{T-T}} ]</td>
</tr>
<tr>
<td>Total Cost Ratio</td>
<td>[ TC_{Ratio} = \frac{n_{T-N} + n_{T-T}}{\lambda \cdot (n_{N-N} + n_{N-T})} ]</td>
</tr>
<tr>
<td>ROC Curve</td>
<td>True positive rate plotted against false positive rate</td>
</tr>
</tbody>
</table>

**Design and Development of Efficient Techniques for Securing E-mail System from threats**

(M. Tariq Banday) Page 156 of 266
Summary

Increasing reliance on the electronic mail (e-mail) has attracted spammers to send more and more spam e-mails in order to maximizing their financial benefits. These unwanted e-mails not only clog the Internet traffic but also cause storage problems at the receiving servers. Besides, spam e-mails also serve as a vehicle to a variety of online crimes and abuses. Spam has also spread to various other services on the Internet including wired and wireless telecommunication services like traditional telephone service, short messaging service (SMS) and multimedia messaging services (MMS) on mobile phones. Though several anti-spam procedures are currently employed to distinguish spam e-mails from the legitimate e-mails but spammers and phishes obfuscate their e-mail content to circumvent anti-spam procedures. Efficiency of anti-spam procedures to combat spam entry into the system greatly depend on their level of operation and a clear insight of various possible modes of spamming. A filter essentially employs some mathematical function for classification, some are adaptive while others are not. The performance of Adaptive filters in terms of various evaluation measures is impressive as they have the ability to learn the changing tactics of the spammers.
Chapter 6: Spam and Anti-Spam

References


Chapter - 6: Spam and Anti-Spam


[6.16] Infra note 6.20 at p 327.


[6.21] See, e.g., LA. REV. STAT. ANN. § 14:73.1(13) (West Supp. 2000) (defining "unsolicited bulk electronic mail" as an e-mail advertisement that is "sent in the same or substantially similar form to more than one thousand recipients"); Act of Apr. 17, 2000, ch. 423, § 1, 2000 Idaho Sess. Laws 1373, 1373


Chapter - 6: Spam and Anti-Spa

[6.31] Rosenberg, Jennings and Peterson 'The session initiated protocol (SIP) and spam' Internet Draft, IETF, Feb, 2007.


[6.40] Lyon and Wong 'Sender ID: Authenticating Email' IETF, RFC 4406, April 2006.


[6.45] Levine, Crocker, Silberman and Finch 'Bounce Address Tag Validation (BATV)' IETF Internet Draft, draft-levine-massbatv-00, September 2004. Also see http://mipassoc.org/batv


[6.55] Dwork and Naor 'Pricing via processing or combating junk mail' In proceedings of CRYPTO'92, (1993) pp. 139-147.


[6.85] Gavrilis, Tsoulos and Dermatas 'Neural Recognition and Genetic Features Selection for Robust detection of E-mail Spam' Lecture Notes in Computer Science, 2006.


