Overview of Anti-spam filtering Techniques

Sushma L.Wakchaure, Shailaja D.Pawar,Ganesh D.Ghuge ,Bipin B.Shinde

Amrutvahini Polytechnic, Sangamner

Professor, Dept. of Computer Engineering, Amrutvahini Polytechnic College, Maharashtra, India.

---***---

Abstract - *Electronic mail (E-mail) is an essential communication tool that has been greatly abused by spammers to disseminate unwanted information (messages) and spread malicious contents to Internet users. Current Internet technologies further accelerated the distribution of spam. Spam-reduction techniques have developed rapidly over the last few years, as spam volumes have increased. We believe that no one anti-spam solution is the "right" answer, and that the best approach is a multifaceted one, combining various forms of filtering with infrastructure changes, financial changes, legal recourse, and more, to provide a stronger barrier to spam than can be achieved with one solution alone. Spam Guru addresses the part of this multi-faceted approach that can be handled by technology on the recipient's side, using plug-in tokenizes and parsers, plug-in classification modules, and machine-learning techniques to achieve high hit rates and low false-positive rates. Effective controls need to be deployed to countermeasure the ever growing spam problem. Machine learning provides better protective mechanisms that are able to control spam. This paper summarizes most common techniques used for anti-spam filtering by analyzing the e-mail content and also looks into machine learning algorithms such as Naïve Bayesian, support vector machine and neural network that have been adopted to detect and control spam. Each machine learning has its own strengths and limitations as such appropriate preprocessing need to be carefully considered to increase the effectiveness of any given machine learning.*

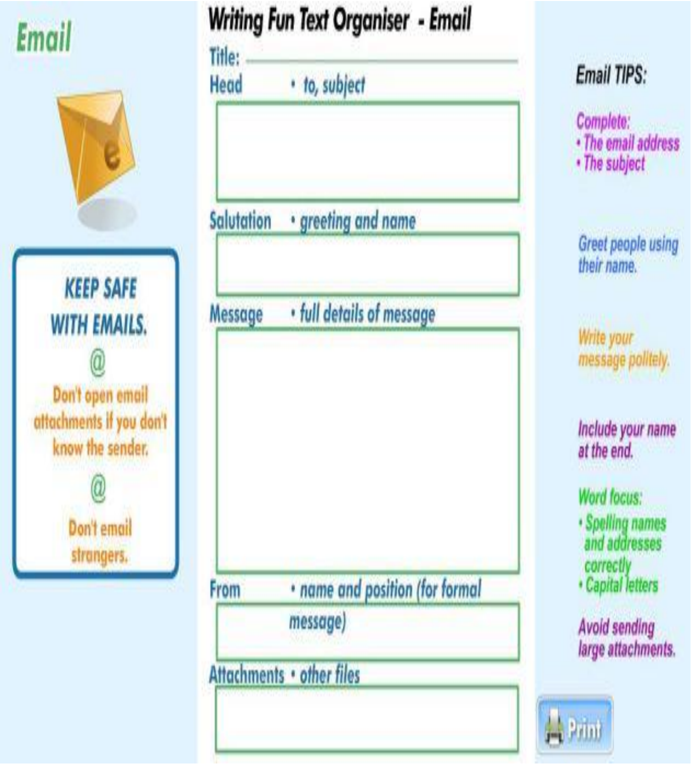
***Key Words*: Anti-spam filters, text categorization, electronic mail (E-mail), Spam Bayes, training email filters, content filtering, false negatives, user level spam filtering machine learning,**

1. INTRODUCTION

Spam-reduction techniques have developed rapidly over the last few years, as spam volumes have increased. We believe that the spam problem requires a multi-faceted solution that combines a broad array of filtering techniques with various infrastructural changes, changes in financial incentives for spammers, legal approaches, and more [1]. This paper describes one part of a more comprehensive anti-spam research effort undertaken by us and our colleagues: SpamGuru, a collaborative anti-spam filter that combines several learning, tokenization, and user interface elements to provide enterprise-wide spam protection with high spam detection rates and low false-positive rates. E-mail or electronic mail is an electronic messaging system that transmits messages across computer networks. Users simply type in the message, add the recipient's e-mail address (es) and click the send button. Users can access any free e-mail service such as Yahoo mail, Gmail, Hotmail, or register with ISPs (Internet Service Providers) in order to obtain an e-mail account at no cost except for the Internet connection charges. Besides that, e-mail can be also received almost immediately by the recipient once it is sent out. E-mail allows users to communicate with each other at a low cost as well as provides an efficient mail delivery system. The reliability, user-friendliness and availability of a wide range of free e-mail services make it most popular and a preferred communication tool. As such, businesses and individual users alike rely heavily on this communication tool to share information and knowledge. Businesses can drastically cut down on communication cost since e-mail is extremely fast and inexpensive; furthermore it is a very powerful marketing tool. Businesses can capitalize from this technology since it is a very popular advertising tool. However, the simplicity of *Corresponding author. E-mail: alaa_taqa@um.edu.my. Sending e-mail and the almost non-existent cost poses another problem: Spam. Spam refers to bulk unsolicited commercial e-mail sent indiscriminately to users. Table 1 enumerates some of them. Based on the Ferris Research (2009), spam can be categorized into the following: 1. Health; such as fake pharmaceuticals; 2. Promotional products; such as fake fashion items (for example, watches); 3. Adult content; such as pornography and prostitution; 4, Financial and refinancing; such as stock kiting, tax solutions, loan packages; 5. Phishing and other fraud; such as "Nigerian 419" and "Spanish Prisoner"; 6. Malware and viruses; Trojan horses attempting to infect your PC with malware; 7. Education; such as online diploma; 8. Marketing; such as direct marketing material, sexual enhancement products; 9. Political; US president votes.

2. E-MAIL STRUCTURE

E-mail messages are divided into 2 parts: Header information and message body. Header information or the header field consists of information about the message's transportation which generally shows the following information; 1. From: displays sender's detail such as e-mail address; 2. To: displays receiver's detail such as e-mail address; 3. Date: displays the date the e-mail was send to the recipient; 4. Received: intermediary server's information and the date the e-mail message is processed; 5. Reply to: reply address; 6. Subject: the subject of message specified by the sender; 7. Message Id: unique id of the message and others The message body contains the message of the e-mail. E-mail messages are presented in plain text or HTML. An e-mail may also have attachments such as graphics, video or other format type and to facilitate these attachments MIME (multipurpose internet mail extension) is used. SPAMMER TRICKS In order to send spam, spammers first obtain e-mail addresses by harvesting addresses through the Internet using specialized software. This software systematically gathers e-mail addresses from discussion groups or websites (Schaub, 2002), other than that spammer also able to purchase or rent collections of e-mail addresses from other spammers or services providers. Table 2 indicates the many tricks used by spammers to avoid detection by spam filters. SPA



Email

KEEP SAFE WITH EMAILS.
@
Don't open email attachments if you don't know the sender.
@
Don't email strangers.

Writing Fun Text Organiser - Email

Title:

Head • to, subject

Salutation • greeting and name

Message • full details of message

From • name and position (for formal message)

Attachments • other files

Email TIPS:

Complete:
• The email address
• The subject

Greet people using their name.

Write your message politely.

Include your name at the end.

Word focus:
• Spelling names and addresses correctly
• Capital letters

Avoid sending large attachments.

Figure 1: Email Structure

3. SPAM CONTROL TECHNIQUES

Anti-spam techniques and methods try to tell apart a spam email from legitimate email. As a typical email consists of few components such as the header, the body and attachments, the algorithms that classify emails may use different features of the mail components to make decision about them. Lot of work has gone into finding solution to spam problem from different dimensions and directions(Islam and Zhou, 2007, Zhang et al., 2012, Xiao-wei and Zhong-feng, 2012, Rajendran and Pandey, 2012, du Toit and Kruger, 2012, Xiao et al., 2010, Wei et al., 2010, So Young and Shin Gak, 2008, Klonowski and Strumiński, 2008, Horie and Neville, 2008, Nhung and Phuong, 2007, McGibney and Botvich, 2007, Liu et al., 2007, Huai-bin et al., 2005, Moon et al., 2004, Wu and Tsai) over the last decade. Various anti-spam solutions are available that have been surveyed by many researchers(Blanzieri and Bryl, 2008, Caruana and Li, 2012, Guzella and Caminhas, 2009, Lai, 2007, Yu and Xu, 2008, Paswan et al., 2012, Nazirova, 2011). Those are blacklists, whitelists, grey lists, content based filtering, feature selection methods, bag-of-words, machine learning techniques such as Naïve Bayes, Support vector machines, artificial neural networks, lazy learning, etc), reputation based techniques, artificial immune systems, protocol based procedures, and so on. (Caruana and Li, 2012) also lists some emerging approaches such as per to peer computing, grid computing, social networks and ontology based semantics along with few other approaches. These solutions can be grouped into various categories such as list based techniques, and filtering techniques; another categorisation can be prevention, detection and reaction techniques(Nakulas et al., 2009). (Paswan et al., 2012) categorises the email spam filtering techniques as origin based spam filtering, content based filtering, feature selection methods, feature extraction methods, and traffic based filtering. The scope of this paper is content based filtering and in specific learning based filters. Hence, we would not go into detail of each of these solutions but limit ourselves to Bayes algorithm. Spammers are insensitive to the consequences of their activities and need to be dissuaded by being made to pay by the internet service providers for the waste of bandwidth occupied by unwanted spam blocked by the servers. This would be a feasible deterrent to reduce spam. To execute this, all service providers must act in CRPIT Volume 149 - Information Security 2014 68 unison and agree to get spammers to pay for spam inconvenience and servers clean up.

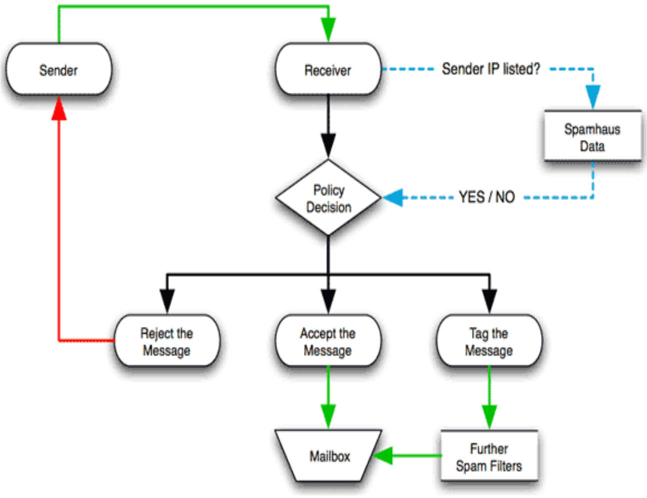
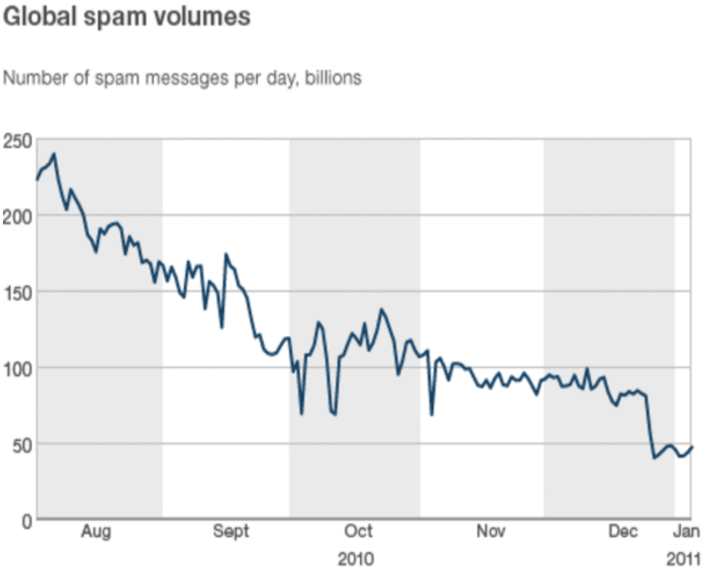


Figure 2: Controlling of Spam mail

4. USER PREFERRED DOMAIN SPECIFIC TRAINING OF FILTER

In this work, we have been able to identity different types of anti–spam techniques exemplified in the use of filters and other characteristic means to deter spammers. Although these anti-spam techniques may be suitable to some users and unsuitable to others, they achieve some level of protection against unwanted email messages. Each of these anti-spam techniques has its unique feature that distinguishes it from the rest of others although none of these are able to perfectly and substantially produce zero false positives and zero false negatives or totally able to stop all real-time and potential spam. The main reason for this is that spammers are always evolving new tricks to deceive the filters. Some well-designed filters (for example, Bayesian Filters) work very well getting success rate as good as 98-99% at certain stages. But this number does not stay the same. Spammers are able to vary these with ease. In this paper, we are verifying that the filtering mechanisms capability can be enhanced by domain specific training and incorporating user preferences. This enhancement of the filtering capability can increase the performance of the filter. We are trying to build over the fact that, there is no correlation between the receiving user's area of interest and content of spam email. Many researchers have studied the content of Spam messages and various categorisations have been published. One categorisation on the basis of content type is Scams, Adult, Financial, Pharmaceuticals, stock, phishing, diploma, software Malware, gambling, dating and so on(SpiderLabs, 2013). Filters are trained to identify spam on the basis of these categories. But such training by the filter would look for features related to any of those categories generally in any email received. The training dataset would contain features from all of these categories and the features could be confusing for the filter as the spam training could still work but the ham features are anything other than these categories. In fact, the cohort of emails in inboxes belonging users in different domains would be different. For examples, user that belongs to healthcare domain would receive healthcare kind of emails as compared to a user who belongs to real estate. Same email may be Ham to one user but Spam to another user based upon their preferences. Users in various Domains have difference preference of emails that they would classify as Spam or Ham. For example, an email from a bookseller trying to sell books on Computer Science would be Spam for a Pharmacist. The interesting question that comes out of this is that how do the filters know which emails are Spam and which are ham for the particular user. Of course in some mail clients such as gmail there is an option of user preference setting where user can be given an option to choose the topic area of interest and then the filter can use that information to classify incoming emails accordingly. There is very little work gone into the area of considering specific user preferences while designing anti-spam filters. (Kim et al., 2007) constructs a user preference ontology on the basis of user profile and user actions and trains the filter on the basis of that ontology. (Kim et al., 2006) suggest user action based adaptive learning where they attach weights to Bayesian classification on the basis of user actions. However none of the work address user belonging to a particular domain and their preferences accordingly. The case of training the filter at the client by the client data is not new; any user who would install SpamBayes would train it with the training data(Meyer and Whateley, 2004). An organisation that uses SpamBayes to filter incoming email would for multiple users would retrain the filter on all received email(Nelson et al., 2009). Training of filter with different feature selection methods is also addressed in (Gomez and Moens, 2010). The novelty in this case is that the data we used for training is the user preferred data carefully collected for a period of 5 years. Second important point that affects the training is that the data collected belongs to a particular user belonging to a specific domain not a general user. This means that we are training the filter that if there is no correlation between the receivers' domain area and the email content, the email is not wanted by the user. User belonging to an educational organisation would have different preferences as compared to a user belonging to a marketing organisation. Different users within an organisation would have different preferences and same message could be classified as spam by different users. The organisational filters cannot take care of such user preferences. Hence, such emails end up in user inboxes as false negatives. The dataset also takes into account such user preferences. The filter is trained on the basis of collection Spam and Ham emails classified by the user belonging to a particular domain. We made a hypothesis that domain specific user preference training of the filter reduces the false negatives in the user inbox. To justify this hypothesis we chose the spam filtering tool called Spam Bayes, installed it on the outlook mail client and trained it with domain specific user preferred data. The next two sections give details on the background of the tool and the experiments done using the datasets.



Source: Symantec

Figure 3: Spam Volume

3. CONCLUSIONS

Email spam has been the focus of studies for a long time. Though there are many different techniques to block spam email messages to reach users inbox, filtering is the most commonly used mechanism and has gained success to some extent. Given the large number of usage of email worldwide, email spam is still plentiful and scale of the problem is enormous. Researchers and organisations make the filers smart and self-learning but spammers are a step ahead. They keep on finding techniques to deceive the filters and their learning mechanisms. Hence, the problem still remains giving scope for researchers to work in the area. This work is an effort in the same scope to reduce false negatives/spam in the inbox of the users which has deceived the organisational filters. It is observed that this further filtering by training the filer with user specific data did make a difference in the amount of false positives. Future work involves creating the feature sets including creating domain specific keywords and list of organisations which can be fed to the filter, conducting experiments and then observing the results to record the improvements. Spam is becoming one of the most annoying and malicious additions to Internet technology. Traditional spam filter software are unable to cope with vast volumes of spam that slip past anti-spam defenses. As spam problems escalate, effective and efficient tools are required to control them. Machine learning approaches have provided researchers with a better way to combat spam. Machine learning has been successfully applied in text classification. Since e-mail contains text, the ML approach can be seamlessly applied to classified spam.

5.EXECUTIVE SUMMARY

- "E-mail is the single most important tool for business communication."
- For many organizations, when their e-mail stops, their ability to conduct business stops too.
- About 80% of the intellectual property of a typical company passes through its e-mail server.
- There's a 72% chance of an e-mail failure in any company each year, lasting an average of 62 hours.
- E-mail is set to grow by 68% in the next 5 years. And legal discovery is a growing consideration - making e-mail management more important.
- The indirect costs of e-mail - mainly loss of productive time - are likely to be overlooked, even though they can be high.
- E-mail costs are virtually impossible to generalize, because one is comparing apples with pears with oranges.
- Stand-alone e-mail is inexpensive and simple. The disadvantages of free services (e.g. Gmail, hotmail) for businesses outweigh the benefit of the small cost savings.
- Collaborative e-mail is typified by Microsoft Exchange, and includes communications features such as shared calendars, tasks and contacts; smartphone and Outlook integration; and central management and storage.
- Collaborative e-mail is relatively expensive.
- Exchange servers can be maintained in-house or housed in a data centre.
- Shared hosted Exchange means renting space on a hosted server that includes many other accounts.
- Cost comparisons for hosted vs in-house Exchange are often published by Exchange hosting providers, and tend to be distorted. Caveat emptor!
- Shared hosted Exchange can be considerably more expensive over time than in-house, which also allows control and flexibility.
- Hosted Exchange (shared or not) is an excellent option for businesses with branches in different geographic locations, and shared hosted Exchange is good for very small companies that can't justify the cost of an in-house server.
- Google Apps is an inexpensive alternative to Exchange, but has several disadvantages and doesn't work as well.

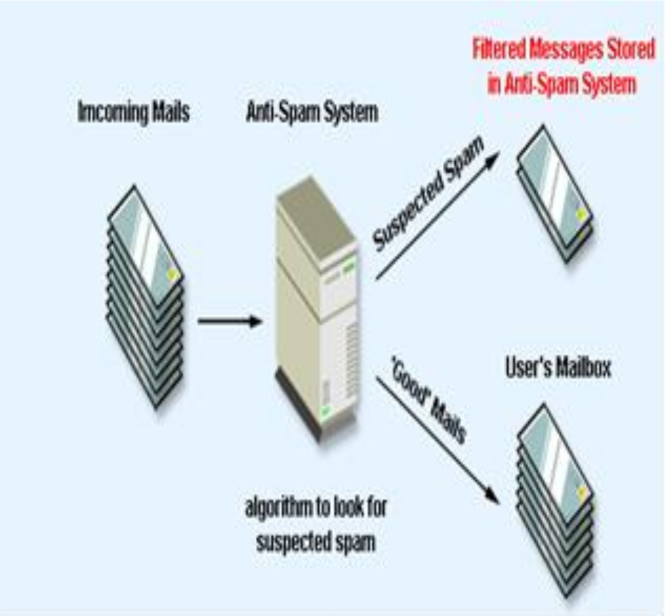

Filtered Messages Stored in Anti-Spam System
Incoming Mails
Anti-Spam System
Suspected Spam
'Good' Mails
User's Mailbox
algorithm to look for suspected spam

Figure 5:Anti-spam filtering system

6. SPAM-FILTERING METHODS

Blacklist

This popular spam-filtering method attempts to stop unwanted email by blocking messages from a preset list of senders that you or your organization's system administrator create. Blacklists are records of email addresses or Internet Protocol (IP) addresses that have been previously used to send spam. When an incoming message arrives, the spam filter checks to see if its IP or email address is on the blacklist; if so, the message is considered spam and rejected.Though blacklists ensure that known spammers cannot reach users' inboxes, they can also misidentify legitimate senders as spammers. These so-called false positives can result if a spammer happens to be sending junk mail from an IP address that is also used by legitimate email users. Also, since many clever spammers routinely switch IP addresses and email addresses to cover their tracks, a blacklist may not immediately catch the newest outbreaks.

Real-Time Blackhole List

This spam-filtering method works almost identically to a traditional blacklist but requires less hands-on maintenance. That's because most real-time blackhole lists are maintained by third parties, who take the time to build comprehensive blacklists on the behalf of their subscribers. Your filter simply has to connect to the third-party system each time an email comes in, to compare the sender's IP address against the list.Since blackhole lists are large and frequently maintained, your organization's IT staff won't have to spend time manually adding new IP addresses to the list, increasing the chances that the filter will catch the newest junk-mail outbreaks. But like blacklists, real-time blackhole lists can also generate false positives if spammers happen to use a legitimate IP address as a conduit for junk mail. Also, since the list is likely to be maintained by a third party, you have less control over what addresses are on — or not on — the list.

Whitelist

A whitelist blocks spam using a system almost exactly opposite to that of a blacklist. Rather than letting you specify which senders to block mail from, a whitelist lets you specify which senders to allow mail from; these addresses are placed on a trusted-users list. Most spam filters let you use a whitelist in addition to another spam-fighting feature as a way to cut down on the number of legitimate messages that accidentally get flagged as spam. However, using a very strict filter that only uses a whitelist would mean that anyone who was not approved would automatically be blocked.Some anti-spam applications use a variation of this system known as an automatic whitelist. In this system, an unknown sender's email address is checked against a database; if they have no history of spamming, their message is sent to the recipient's inbox and they are added to the whitelist.

Greylist

A relatively new spam-filtering technique, greylists take advantage of the fact that many spammers only attempt to send a batch of junk mail once. Under the greylist system, the receiving mail server initially rejects messages from unknown users and sends a failure message to the originating server. If the mail server attempts to send the message a second time — a step most legitimate servers will take — the greylist assumes the message is not spam and lets it proceed to the recipient's inbox. At this point, the greylist filter will add the recipient's email or IP address to a list of allowed senders.Though greylist filters require fewer system resources than some other types of spam filters, they also may delay mail delivery, which could be inconvenient when you are expecting time-sensitive messages.

Content-Based Filters

Rather than enforcing across-the-board policies for all messages from a particular email or IP address, content-based filters evaluate words or phrases found in each individual message to determine whether an email is spam or legitimate.

Word-Based Filters

A word-based spam filter is the simplest type of content-based filter. Generally speaking, word-based filters simply block any email that contains certain terms.Since many spam messages contain terms not often found in personal or business communications, word filters can be a simple yet capable technique for fighting junk email. However, if configured to block messages containing more common words, these types of filters may generate false positives. For instance, if the filter has been set to stop all messages containing the word "discount," emails from legitimate senders offering your nonprofit hardware or software at a reduced price may not reach their destination. Also note that since spammers often purposefully misspell keywords in order to evade word-based filters, your IT staff will need to make time to routinely update the filter's list of blocked words.

Heuristic Filters

Heuristic (or rule-based) filters take things a step beyond simple word-based filters. Rather than blocking messages that contain a suspicious word, heuristic filters take multiple terms found in an email into consideration.Heuristic filters scan the contents of incoming emails and assigning points to words or phrases. Suspicious words that are commonly found in spam messages, such as "Rolex" or "Viagra," receive higher points, while terms frequently found in normal emails receive lower scores. The filter then adds up all the points and calculates a total score. If the message receives a certain score or higher (determined by the anti-spam application's administrator), the filter identifies it as spam and blocks it. Messages that score lower than the target number are delivered to the user.Heuristic filters work fast — minimizing email delay — and are quite effective as soon as they have been installed and configured. However, heuristic filters configured to be aggressive may generate false positives if a legitimate contact happens to send an email containing a certain combination of words. Similarly, some savvy spammers might learn which words to avoid including, thereby fooling the heuristic filter into believing they are benign senders.

Bayesian Filters

Bayesian filters, considered the most advanced form of content-based filtering, employ the laws of mathematical probability to determine which messages are legitimate and which are spam. In order for a Bayesian filter to effectively block spam, the end user must initially "train" it by manually flagging each message as either junk or legitimate. Over time, the filter takes words and phrases found in legitimate emails and adds them to a list; it does the same with terms found in spam.To determine which incoming messages are classified as spam, the Bayesian filter scans the contents of the email and then compares the text against its two-word lists to calculate the probability that the message is spam. For instance, if the word "valium" has appeared 62 times in spam messages list but only three times in legitimate emails, there is a 95 percent chance that an incoming email containing the word "valium" is junk.Because a Bayesian filter is constantly building its word list based on the messages that an individual user receives, it theoretically becomes more effective the longer it's used. However, since this method does require a training period before it starts working well, you will need to exercise patience and will probably have to manually delete a few junk messages, at least at first.

ACKNOWLEDGEMENT

I would like to give my sincere thanks to my thanks go to all the people who have supported me to complete the research work directly or indirectly. I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future.

REFERENCES

[1] B. Leiba and N. Borenstein. A Multi-Faceted Approach to Spam Prevention, Proceedings of the First Conference on E-mail and Anti-Spam, 2004.

[2] M. Leonard, M. Rodriguez, R. Segal, and R. Shoop. Managing Customer Opt-Outs in a Complex Global Environment. Proceedings of the First Conference on E-mail and Anti-Spam, 2004.

[3]P. Graham, A Plan for Spam. http://paulgraham.com/spam.html, August 2002.

[4]P. Graham, Better Bayesian Filtering. http://paulgraham.com/better.html, January 2003.

[5] P. Capek, B. Leiba, and M. N. Wegman. Charity Begins at your Mail Server, http://www.research.ibm.com/people/w/wegman/charity.htm, 2004.

[6] J. Lyon and M. Wong. MTA Authentication Records in DNS, Internet Draft, http://www.ietf.org/internetdrafts/draft-ietf-marid-core-01.txt, June 2004

[7] S. Schleimer, D. Wilkerson, and A. Aiken, Winnowing: local algorithms for document fingerprinting. In Proceedings of SIGMOD 2003, San Diego CA, June 9-12, 2003.

[8] T. Zhang and F. J. Oles. Text categorization based on regularized linear classification methods. Information Retrieval, 4:5-31, 2001.

[9] CLIFFORD, M., FAIGIN, D., BISHOP, M. & BRUTCH, T. Miracle Cures and Toner Cartridges: Finding Solutions to the Spam Problem. 19th annual computer security applications conference (ACSAC 2003), 2003.

[10] DEEPAK, P. & SANDEEP, P. Spam filtering using spam mail communities. In: SANDEEP, P., ed. Applications and the Internet, 2005. Proceedings. The 2005 Symposium on, 2005.

[11] NAGAMALAI, D., DHINAKARAN, C. & LEE, J. K. Multi Layer Approach to Defend DDoS Attacks Caused by Spam. In: DHINAKARAN, C., ed. Multimedia and Ubiquitous Engineering, 2007. MUE '07. International Conference on, 2007.

[12] Ferris Reseacrh (2009). Spam, Spammers, and Spam Control A White Paper by Ferris Research (March 2009).

[13] Pang X, Feng Y, Jiang W (2007). A Chinese Anti-Spam Filter Approach Based on Support Vector Machine. Management Science and Engineering, 2007. ICMSE 2007. International Conference on, 20-22: 97-102, Aug. 2007.

[14] Rish I (2001). An empirical study of the naïve bayes classifier

[15] JUNG, J. & EMIL, S. An Empirical Study of Spam Traffic and the Use of DNS Black Lists. Internet Measurement Conference,, October 2004 Taormina, Italy.

[16] G. Sakkis, I. Androutsopoulos, G. Paliouras, V. Karkaletsis, C. D. Spyropoulos and P. Stamatopoulos, Stacking Classifiers for Anti-Spam Filtering of E-Mail, In Proceedings of EMNLP-01, 6th Conference on Empirical Methods in Natural Language Processing, 2001.

[17] P. Pantel and D. Lin. SpamCop -- A Spam Classification & Organization Program. In Proceedings of AAAI-98 Workshop on Learning for Text Categorization, 1998.

[18] I. Rigoutsos and T. Huynh. Chung-Kwei: a pattern-discovery-based system for the automatic identification of unsolicited e-mail messages (SPAM). Proceedings of the First Conference on E-mail and Anti-Spam, 2004.