A MULTI-LAYER ARCHITECTURE FOR SPAM-DETECTION SYSTEM

Vivek Shandilya, Fahad Polash and Sajjan Shiva

Department of Computer Science, University of Memphis, Memphis, TN, USA
vmshndly,fpolash,sshiva@memphis.edu

ABSTRACT

As the email is becoming a prominent mode of communication so are the attempts to misuse it to take undue advantage of its low cost and high reachability. However, as email communication is very cheap, spammers are taking advantage of it for advertising their products, for committing cybercrimes. So, researchers are working hard to combat with the spammers. Many spam detections techniques and systems are built to fight spammers. But the spammers are continuously finding new ways to defeat the existing filters. This paper describes the existing spam filters techniques and proposes a multi-level architecture for spam email detection. We present the analysis of the architecture to prove the effectiveness of the architecture.

KEYWORDS

Email, Spam, Spam detection, Filters, Multi-Layer Architecture

1. INTRODUCTION

In general, unsolicited emails are regarded as spam email. But according to Mail Abuse Prevention System, L.C.C. [2], the three conditions to consider an email as spam are: 1) The recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients, 2) The recipient has not verifiably granted deliberate, explicit, and still revocable permission for it to be sent, and 3) The transmission and reception of the message appears to the recipient to give a disproportionate benefit to the sender. Spam affects the users in several ways. The user will lose productive time by looking into the spam emails. The user mailbox is overburdened by spam emails. Spam emails consume network bandwidth. Radicati Research Group Inc., a Palo Alto, CA, based research firm, estimates that spam costs businesses $20.5 billion annually in decreased productivity as well as in technical expenses. Nucleus Research estimates that the average loss per employee annually because of spam is approximately $1934[3]. The main success of the spammer is to sell a product advertised in the spam email. Though most of the users ignore the advertisement, even if some order the advertised product, it is profitable for the spammer, as it costs very less to send millions of spam. An internet connection and a single click is enough for the spammer to send a spam email to many email users. According to the industry figures, 1 out of the 12.500.000 spam messages that are sent, lead to a sale [4]. Spammers get a high percentage profit share for each of the sale generated.

The damage due to spam has met with many attempts to detect and stop them. Many commercial spam email filters are available in the market [5]. For example some of the client side filters are: ASB AntiSpam, Outlook Spam Filter, Spam Alarm, SpamButcher, Qurb Spam, Spam Arrest, Spam Bully, MailWasher Pro, McAfee SpamKiller, Feox for Outlook/OE, Edovia AntiSpam, SAproxy Pro, Dewqs' NMS for Outlook, AntiSpamWare and LashBack. Some of the server side

Dhinaharan Nagamalai et al. (Eds) : ACITY, WiMoN, CSIA, AIAA, DPPR, NECO, InWeS - 2014
pp. 193–200, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4520

spam filters are: GFI Anti Spam Filter,M-Switch Anti-Spam, Astaro Security Gateway, Hexamail Guard, Symantec AntiSpam for SMTP, Accessio Server, SpamSentinel for Domino Server and Kaspersky Anti-Spam Enterprise Edition by Alligate. In spite of active countermeasures spamming is thriving. In November 2013, the percentage of spam email was 72.5% out of all email traffic [1]. This calls for continuous efforts to discourage the spammers.

In this paper, we propose a multi-level architecture which will combine the existing spam filters in different layers. This architecture could be used as a generic framework for spam email detection. Existing techniques employed at each layer are also described. Our main contribution can be summarized as,

1. A multi-layer architecture using the present day state of art spam detection technologies.
2. Analysis to prove the advantages of our novel method of having two thresholds over the traditional single threshold based classification, in improving the accuracy and reducing the false positives while keeping the computational load for filtering low when using each of the filtering features.

The rest of the paper is organized as follows: Section 2 describes related works regarding multi-level spam detection approaches, Section 3 describes the proposed architecture, Section 4 presents the performance evaluation measurements of spam detection and we conclude in Section 5.

2. RELATED WORKS

Many researchers have already given efforts to fight with the spammers. Some of the works are related to our proposed multi-level architecture for spam detection. However, our proposed model is different from these salient works. Jianying et al. [6] describe a multi layer framework for spam detection. They divide the spam detection techniques between server and client side deployments. Our proposed model does not differentiate between server and client. It can be equally applied to both server and client side anti-spam countermeasures. Rafiul et al. [7] proposes a multi-tier classification for phishing email. The classification result in the first tier is given to a second tier classifier. If the classification of the second tier and first tier matches, then the result is considered as the right output. But if the results differ, then a third tier classifier is used to classify the email. The output of the third classifier is considered the correct classification of the email. Thus, best of three classifiers are used to get the classification of an email. But in our proposed model, if any layer can classify an email with the confidence above the threshold, then the lower level is not invoked. And our proposed model can handle more than three levels of classifier to reduce the false positive rate as much as possible. In [8], Xiao et al. proposed a hierarchical framework for spam email detection. The first layer in their framework is a text classifier. But in our case, we have considered other behavioural features of spam email like blacklisting sender, sender reputation etc. Again, we have included negative selection based detection in the last layer which will be more effective against new spam emails. Zhe et al. [9] presents an approach targeting mainly at image spam email. The architecture presented by them is two layered. The first layer classifies non image spam and the second layer classifies image spam. The second layer involves multiple spam filters which will take longer time for training the detectors. Our proposed multi-level architecture is presented in Figure 1. The purpose of this model is to put the existing techniques in an organized way so that the detection of the spam would be faster and the false positive rate would be lower. If the detection of the spam could be possible in the upper layer, then the lower layer would not be invoked, and thus it will reduce the computational load. And in each layer, we can increase the threshold value so that the rate of the false positive would be minimum. As a result the overall performance of the spam filter detection process would be better.

This model comprises of the following layers: 1) Blacklist/Whitelist layer, 2) Content based filter, 3) Image based filter, 4) Negative Selection of Unknown Spam Email, and 5) Recipient Decision. The description of each layer is given below.

3. MULTI-LAYERED ARCHITECTURE

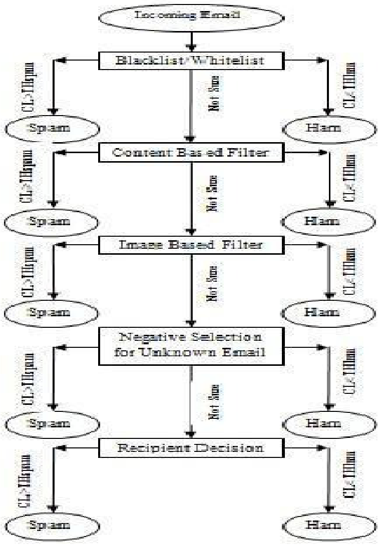


Figure 1. Multi-Layer spam detection Architecture

Our proposed multi-level architecture is presented in Figure 1. This model organizes the existing techniques for better detection of spam with lower false positive rate. If the detection of the spam could be possible in the upper layer, then the lower layer would not be invoked, and hence reduces computational expenditure. And in each layer, we can increase the threshold value so that the rate of the false positive would be minimum. As a result the overall performance of the spam filter detection process would be better. This model comprises of the following layers: 10 Black list/White list Layer, 20. Content based Filter, 3). Image Based Filter, 4). Negative Selection of Spam email and 5). Recipient decision. The description of each layer is given below.

3.1. Blacklist / Whitelist

Blacklist and Whitelist filters can classify the emails without reading the messages. Based on the senders reputations, blacklist and whitelist are prepared. Blacklisted senders' emails are classified as spam whereas whitelisted senders' emails are classified as ham emails. Any user can add email sender's email address(and in advanced cases IP addresses) in the blacklist or whitelist. The advantage of this approach is that the classification is very fast as it does not require to go through the messages. However, one disadvantage of this approach is that it requires the blacklist/whitelist to be updated regularly. Otherwise the false positive rate would be higher.

Duncan et al. [10] proposes a mechanism by which the blacklist will be updated dynamically. By checking the log files of a particular IP address and then other suspicious activities of the sender, the sender’s IP address is added in the blacklist. Anirudh et al. [11] proposes a technique with which the blacklist is updated based on how the sender is sending email, rather than not relying on IP address, the sending pattern is observed and the IP address of the sender is added in the blacklist.

3.2. Content Based Filter

Content based filter requires the whole message to be read before taking the decision. As a result it will take computationally more time than the Black list /White List layer. Machine learning and fingerprint based filters are popular among the content based filters.

Fingerprint filters generate unique fingerprints for known spam messages and store in a database. It compares the fingerprint of the incoming email to that of stored spam messages. If a match is found for the spam messages' fingerprint, the email is classified as a spam. However the matching should be above a certain threshold level. Damiani et al. [12] proposes a robust way to generate the fingerprint of an email.

There are several methodologies in machine learning techniques. Statistical and artificial immune systems are notable among them. Support Vector Machine (SVM) is a famous statistical tool. In SVM, each email is considered as an n-dimensional vector. Each dimension could be the frequency of a certain word. Harris et al. [13] has compared among different algorithms for statistical filtering. The results of their experiments proves that SVM is better than Ripper, Rocchio, and Boosting Decision Trees algorithm. Mehran et al. [14] presented bayesian classification for spam email where some domain specific features like the domain of the sender(.edu or .gov), the time of sending the email, whether the email contains attachment are taken into consideration. Their experiments show that bayesian classifier works better when the classification is done with additional domain specific features along with the contents of the message. Dat et al. [15] have proposed an approach to detect misspelled spam words in spam email. For example, the word 'viagra' is commonly used in spam emails. So, this word is a blacklisted word. To defeat the spam filter, spammer can use the word 'viaaagra'. The possibility theory will calculate all the possibilities of misspelling of spam keyword and thus can classify the email accurately. Clotilde et al. [16] presents a symbiotic filtering approach where trained filters are exchanged among the users. As it exchange filters, not emails, so the communication and computational cost is minimum in this approach while it achieves better performance.

Artificial Immune System uses machine learning methods inspired by the human immune systems for fighting the spam. Human immune system distinguishes between self and non-self, and artificial immune system distinguishes between a self of legitimate email and a non-self of spam email. The heart of artificial immune system is detectors which are randomly generated from a set of gene library. Oda et al. [17] proposes the approach of artificial immune system in spam detection successfully.

3.3. Image Based Filter

As text based filters can classify emails successfully, spammers are taking resort to image spam emails in order to defeat the existing text based filters. Initially, the optical character recognition(OCR) was being used to detect the text embedded in the image. However, spammers use randomization in creating image spam to defeat OCR. For example, spammers introduce additional dots, frames, bars in the image. They can change the font type of the text included in the image. Zhe et al. [9] proposes a technique which is effective in image spam detection. They have involved three different types of image filters: Color Histogram Filter, Haar Wavelet Filter and Orientation Histogram Feature. Each of the filters works better in different types of randomization detection. After combining the output from three different filters, decision is taken to classify the email. Their experiments have shown less than 0.001% false positive rate in image spam detection. Uemera et al. [18] proposes an image filtering technique based on Bayesian filter. The filter will consider the image file size, file name, compressibility technique and area of the image to classify the spam image email. Their experiment also exhibited low false positive rate.

3.4. Negative Selection of Unknown Spam Email

Negative selection is a part of Artificial Immune System. Dat et al. [19] proposed a novel technique for spam email detection based on negative selection. The difference between this filter and others is that negative selection does not require any prior knowledge of spam emails. It does not require any prior training. As a result, this filter can be used readily. And as such the unknown spam emails could be classified by the technique proposed by Dat.

3.5. Recipient Decision

If the above layers cannot classify the incoming email either as ham or spam with confidence level above the required threshold, then the email could be tagged with a probability number of being a ham or spam, but not classified and let into the inbox. The user can decide whether the email should be forwarded to spam or inbox folder. Based on the decision of the user, the filters of the upper layer can learn and use the knowledge for future classification of this sort of email. Elena et al. [20] proposed a spam filtering technique based on the reputation of the reporters. Whenever any of the users report any email as spam, the system maintains the trustworthiness of the reporter and use the feedback to classify emails.

4. PERFORMANCE ANALYSIS

In spam email detection, if a spam is detected correctly, it is called true positive (TP), if a legitimate email is classified correctly, it is called true negative (TN). Similarly, the misclassification of legitimate email into spam email is called false positive (FP) and the misclassification of spam email as a legitimate email is called false negative (FN). The goal of the spam detection is to classify as many as possible emails correctly and at the same time to reduce the false positive rate. Because if any legitimate email is classified as spam and the user overlooks that email, he/she might miss valuable information. As a result the cost of a false positive classification is very high. Description of various parameters [21] for spam detection are given below:

a) Recall = TP/ (TP + FN). It explains how good a test is at detecting the positives. i.e. predicting positive observations as positive. A high recall is desired for a good model. Recall is also known as sensitivity or TP Rate.

b) FP Rate = FP/ (FP+ TN). It explains how good a model is at detecting the negatives. A low value is desirable.

c) Precision = TP/ (TP + FP). It determines how many of the positively classified are relevant. It is the percentage of positive classifications being correct. A high precision is desirable.

d) Accuracy = TP + TN/ (TP+TN+FP+FN). It tells how well a binary classification test correctly i.e. what percentage of predictions that are correct. Accuracy alone is not a good indicator, as it does not tell how well the model is in detecting positives or negatives separately.

In our proposed model the classification of the emails would be done as follows:

Each filter calculates the correlation factor of the incoming mail based on the known characteristics of the spam based on the features of the filter. Then the correlation factor is compared to two threshold values, which are calibrated to decide if the email is a spam, a ham or not decidable with the information at hand. If the email is classified into the third category, it is sent to another more rigorous and computationally expensive filtering layer. This architecture leads us to have a classification which is having an acceptable accuracy and an acceptable false-positive rate, while considering each of the features when situation requires. As we show in the analysis below with the given technologies at hand this architecture provides an improvement over using the filters individually as some of the features considered by one filter will not be considered by the other filter.

Let A= Accuracy, T = Percentage of Correctly Classified emails, F= Percentage of Incorrectly classified emails = False positives + False Negatives. Then by definition we have, A = 1/(1+(F/T)). The effectiveness of the filtering depends on the nature of the incoming traffic and the filter's response to it. More precisely, when a filter can expect that, in the incoming set of emails, if there are spams, and those spams have a particular feature, then it can check for high correlation for that feature and classify successfully the spam emails and keep them out of inboxes of the end users. The only thing that the filters can control is the threshold of correlation factor to classify the incoming email as spam or ham. If the threshold for classification as spam is high then many spam end up in the inbox. If it is low, then many hams may be misclassified as spam, increasing the false positive. It is learnt from experience that false positive should be as less as possible even if that allows some spams to enter the inbox.

Thus let us have $\alpha_{1,}$ $\alpha_{2,}$ α_3 and α_4 as the percentage of incoming emails expected to be classified as spam emails and $\beta_{1,}$ $\beta_{2,}$ β_3 and β_4 be the expected to be classified as ham emails respectively at each layer. Let and γ_1, $\gamma_{2,}$ γ_3 and γ_4 be the percentage of false positives and δ_1 , $\delta_{2,}$ δ_3 and δ_4 be the false negatives of each layer correspondingly from top layer to the bottom layer in Figure 1. Then we get the total False Positive = $\alpha_1\gamma_1 + (1-(\alpha_1 + \beta_1)) \alpha_2 \gamma_2 + (1-(\alpha_1+ \alpha_2+ \beta_1+ \beta_2)) \alpha_3 \gamma_3 + (1-(\alpha_1+ \alpha_2+ \alpha_3 + \beta_1+ \beta_2 + \beta_3) \alpha_4 \gamma_4$. Similarly we have the total False Negative = $\beta_1 \delta_1 + (1-(\alpha_1 + \beta_1)) \beta_2 \delta_2 + (1-(\alpha_1+ \alpha_2+ \beta_1+ \beta_2)) \beta_3 \delta_3 + (1-(\alpha_1+ \alpha_2+ \alpha_3 + \beta_1+ \beta_2 + \beta_3) \beta_4 \delta_4$. On simplification we can verify that the percentage of emails the effective false positives = $(\alpha_1\gamma_1 + \alpha_2 \gamma_2 + \alpha_3 \gamma_3 + \alpha_4 \gamma_4)$ – (non-negative term) = Sum of False positives of individual filters – positive term. So, we know that collectively the false positive and by symmetry false negative are better off than if we had considered individual filters separately.

5. CONCLUSIONS

In this paper we have proposed a multi-layer architecture which provides a layered approach for spam detection process using the existing techniques. As the spammers are coming up with new ways to defeat the existing filters, continuous efforts are required to improve the filters in each layer. Detecting spam email closer to the source will avoid wasting bandwidth and traffic processing. With our analysis we show that our architecture yields more correct classifications for the same given thresholds of spam without adversely affecting false positives and the threshold of ham affecting the false negatives. Further research into the exact machine learning algorithms to detect spam incorporating this overall architecture for the layers would lead to better preparedness to fight the increasing spam traffic.

REFERENCES

[1] http://www.securelist.com/en/analysis/204792321/Spam_in_November_2013

[2] http://www.sans.org/reading-room/whitepapers/email/is-affect-us-deal-spam-1111

[3] http://www.spamlaws.com/spam-stats.html

[4] http://www.spamexperts.com/en/news/motivation-spammers

[5] http://www.spamhelp.org/software/

[6] Jianying Zhou, Wee-Yung Chin, Rodrigo Roman, and Javier Lopez,(2007) "An Effective Multi-Layered Defense Framework against Spam", Information Security Technical Report 01/2007.

[7] Rafiqul Islam,JemalAbawajy,"A multi-tier phishing detection and filtering approach (2013)",Journal of Network and Computer Applications,Volume 36, Issue 1, January, pp. 324–335.

[8] Xiao Mang Li, Ung Mo Kim,(2012)"A hierarchical framework for content-based image spam filtering", 8th International Conference on Information Science and Digital Content Technology (ICIDT), Jeju,June , pp. 149-155.

[9] Z. Wang, W. Josephson, Q. Lv, M. Charikar and K. Li.(2007) Filtering Image Spam with near-Duplicate Detection, in Proceedings of the 4th Conference on Email and Anti-Spam CEAS.

[10] Duncan Cook, Jacky Hartnett, Kevin Manderson and Joel Scanlan(2006),"Catching spam before it arrives: domain specific dynamic blacklists", ACSW Frontiers '06 Proceedings of the 2006 Australasian workshops on Grid computing and e-research - Volume 54, January , pp. 193-202.

[11] Anirudh Ramachandran, Nick Feamster, and Santosh Vempala,(2007)"Filtering spam with behavioral blacklisting", Proceedings of the 14th ACM conference on Computer and communications security CCS'07, NY, USA, pp. 342-351.

[12] Damiani E., Vimercati S. D. C. d. et al.,(2004) "An Open Digest-based Technique for Spam Detection", San Francisco, CA, USA, pp. 1-6.

[13] Harris Drucker,Donghui Wu, and Vladimir N. Vapnik,(1999)"Support Vector Machines for Spam Categorization", IEEE Transactions On Neural Networks, Vol. 10, No. 5, September.

[14] Mehran Sahami, Susan Dumais, David Heckerman, and Eric Horvitz(1998),"A Bayesian Approach to Filtering Junk E-Mail",Learning for Text Categorization: Papers from the 1998 Workshop, Madison, Wisconsin, AAAI Technical Report WS-98-05.

[15] Dat Tran, Wanli Ma, Dharmendra Sharma, and Thien Nguyen,(2007)"Possibility Theory-Based Approach to Spam Email Detection",IEEE International Conference on Granular Computing.

[16] Clotilde Lopes, Paulo Cortez, Pedro Sousa, Miguel Rocha, and Miguel Rio,(2011) “Symbiotic filtering for spam email detection”, Expert Systems with Applications: An International Journal, Volume 38 Issue 8, August, pp.9365-9372.

[17] Oda, T. and T. White.(2005) Immunity from Spam: An Analysis of an Artificial Immune System for Junk Email Detection. in 4th International Conference on Artificial Immune Systems (ICARIS).

[18] M. Uemura and T. Tabata,(2008)"Design and Evaluation of a Bayesian-filter based Image Spam Filtering Technique", in Proceedings of the International Conference on Information Security and Assurance(ISA).

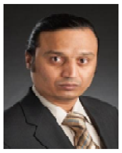
[19] Dat Tran, Wanli Ma, and Dharmendra Sharma,(2009) "A Novel Spam Email Detection System Based on Negative Selection",In Proceedings of the 4th International Conference on Computer Sciences and Convergence Information Technology (ICCIT’09). Los Alamitos, CA, 2009,pp. 987-992.

[20] Elena Zheleva, Aleksander Kolcz and Lise Getoor,(2008) "Trusting spam reporters: A reporter-based reputation system for email filtering", ACM Transactions on Information Systems (TOIS),Volume 27 Issue 1, December.

[21] P.K Panigrahi,(2012)"A Comparative Study of Supervised Machine Learning Techniques for Spam E-mail Filtering", Fourth International Conference on Computational Intelligence and Communication Networks (CICN), Mathura, Nov,pp.506-512.

AUTHORS

Vivek Shandilya holds a BE in Electronics and Communication Engineering from Bangalore university, MS in Computer Science and is a PhD candidate in Computer Science at University of Memphis. His research areas are optimization and security of stochastic systems.



Fahad Polash is a PhD student at Department of Computer Science at university of Memphis. He has a bachelor's degree in computer science and worked in telecom industry before starting his graduate studies. His research areas are computer networking, network security and forensics.

Sajjan Shiva is the chair and a professor at Department of Computer Science at university of Memphis. He was formerly chair of the Department of Computer Science at University of Alabama, Huntsville. His research areas are computer organization and architecture, parallel processing, software engineering, security systems and cloud computing. He is a fellow of IEEE