Different Techniques for Spam Filtering: A Survey
Vandana S. Ahire

*Master Student, Department of Computer Engineering, S.S.V.P.S’s B.S. Deore College of Engineering, Dhule, Maharashtra, India.

ABSTRACT:
Spam is currently of grave and escalating concern and it is challenging to develop spam filters. There are many different techniques to develop a correct and user-friendly spam filter. In this paper, the overview of existing e-mail spam filtering methods like keyword based, Machine learning, Neural networks is given. Recently spam filtering is also done using Social network called as SOAP. It exploits the social relationships among email correspondents and their (dis)interests to detect spam adaptively and automatically.

Keywords: Spam filter, social networks, Bayesian spam filters

I. INTRODUCTION
The email spam consume a large number of server resources and bring great harms to the user's normal use. Spam is unsolicited email message which is forced on people who would not otherwise choose to receive it. Spam is a combination of: unsolicited commercial e-mail (UCE) and unsolicited bulk e-mail (UBE)[6]. The three main characteristics of spam email are; (i) not requested by the recipients; (ii) has commercial value; (iii) always sent in bulk. Spam is a problem because it wastes the email storage, email recipients time to open, read and delete the email. It targets individual users with direct mail messages[6].

To protect against unsolicited emails, there are many recent research presented with goal of efficient, accurate spam filtering. Few previous spam filters can meet the requirements of being user-friendly, attack-resilient, and personalized. Most approaches do not take into account the closeness relationships and (dis)interests of individuals. Previous spam filtering approaches can be mainly divided into two categories: content-based and identity-based[1].

II. LITERATURE SURVEY
Spam filtering is used to detect unsolicited and unwanted email and prevent those incoming messages to a user's inbox. There are many spam filtering techniques, in which the email is gauged after it arrives in the Mailbox.

A. Content-Based Approaches
Content-based filtering approaches are based on the assumption and it reads the text in order to discover distinctive features which are used for classifying a message. It is used to analyses the headers, subject, and body of an email message using feature (token) matching or statistic methods to determine whether it is spam or legitimate email. Many content-based filtering methods have been proposed to filter spam from email. An important part in content based filtering is feature selection[1].

The basic approach of content-based spam filtering is the static keyword list [1], which however makes it easy for a spammer to evade filtering by tweaking the message. The second category approaches includes machine learning-based approaches such as Bayesian filters [2], decision trees [9], Support Vector Machines [9], Bayes Classifiers [2], [3], [4] and combinations of these techniques [1].

A. Keyword Based:
It is another approach for spam email filtering with effective results and is commonly employed in commercial software. Here a list of spam words is maintained and the incoming emails are matched for those words. Emails containing the Spam words are classified as Spam[1].

B. Decision Tree
Decision Trees (DT) tree learning algorithms work based on processing and deciding upon attributes of the data. Attributes in DT are nodes and each leaf node is representing a classification. The disadvantages of DT are focus on continues attributes, computational efficiently with growing tree size[8].

C. Machine Learning Classification Approaches:
i. Bayesian Classification:
Bayesian classification theory was derived from the Bayes's theorem in probability theory. Bayes's Theorem is a mathematical formula used for calculating conditional probabilities. Probability-based Bayesian filtering technique is an advanced keyword filtering algorithm. It does not require pre-set rules and analysis of message content. If the calculated probability value is higher than the preset threshold, the message is classified as a spam, and treated accordingly [1].

A Bayesian filter has a list of keywords along with their probabilities to identify an email as a spam email or a legitimate email. The list is built by training the filter. During training, a user is given a pool of emails, and s/he will manually indicate whether each email is spam or not. We use \( P(S) \) and \( P(L) \) to denote the probability that an email is a spam email and a legitimate email, respectively. The filter parses each email for spam keywords. It calculates the probabilities that a word \( w \) appears in a spam email and in a legitimate email, denoted by \( P(w|S) \) and \( P(w|L) \) respectively.

After training, the calculated probabilities are used to compute the probability that an email with a particular set of keywords in it belongs to either category [1], [2], [3]. The probability that an email including a word \( w \) is spam is:

\[
P(S|w) = \frac{P(w|S)P(S)}{P(w|S)P(S) + P(w|L)P(L)}
\]

In Bayesian Spam Filtering [1], [2], [3], following steps are there:
1. Create database of Spam and Non Spam messages.
2. Calculate appearance rate for each independent word (in each Database).
3. Create hash tables for each. These store token to appearance rate.
4. For each incoming mail, calculate the probability of word in both hash tables.
5. Create a new hash table to store this token.
6. Use Bayes formula to calculate the probability of message being spam.
7. If the value is higher than threshold, its SPAM, otherwise it is legal mail.

ii. **Neural Network Method:**

Artificial Neural Networks (ANN) are one of the common classification methods in data mining [8], [9]. Neural Network based classifiers, Multi Layer Perceptron (MLP) and Radial Base Function (RBF) were used in this work. MLP is a feed forward network that makes a model to map input data to output data. Hidden layer in MLP can include various layers between input and output. RBF is another type on ANN. The input of NN in RBF is linear and the output is nonlinear. The output of this type of ANN is taken from weighted sum of hidden layer’s output. The RBF networks are divided in two feed-forward layer [8].

D. **Identity-Based Approaches**

A very different approach to spam detection is based on the behavior of email senders. The simplest identity-based spam filtering approaches are blacklist and white list, which check the email senders for spam detection. A blacklist is a list of senders whose emails are blocked from getting through to the recipients. A whitelist is just the exact opposite. While a blacklist specifies who is to be kept out allowing all others to pass, a whitelist only allows those who are already on the list to get through. Since spammers almost always spoof the “From" field of spam messages, blacklists usually keep IP addresses rather than email addresses. For incoming messages from senders not on the lists, content-based filters may be applied so that the two approaches can complement each other [5]. White lists and blacklists both maintain a list of addresses of people whose emails should not and should be blocked by the spam filter, respectively. One server side solution records the number and frequency of the same email sent to multiple destinations from specific IP addresses. If the number and frequency exceed thresholds, the node with the specific IP address is blocked [1], [5].

Boykin et al. [1], constructed a graph in which vertices represent email addresses and direct edges represent email interactions. Emails are identified as spam, valid, or unknown based on the local clustering coefficient of the graph subcomponent. This is based on the rationale that the social communication network of a normal node has a higher clustering coefficient than that of a spam node. RE is a white list spam filtering system based on social links. It is based on the assumption that all friends and FoF are trustable. Hameed [1] has LENS, which extends the FoF network by adding trusted users from outside of the FoF networks to mitigate spam beyond social circles.

Only emails to a recipient that have been touched by the trusted nodes can be sent into the network. DeBarr et al. [1] evaluated the use of social network analysis measures to improve the performance of a content filtering model. They tried to detect spam by measuring the degree centrality of message relay agents and the average path length between senders and receivers. They claimed that the messages from a promiscuous mail relay or messages with unusual path lengths that deviate from the average are more likely to be spam. Tran et al. [1] implemented an email client called Social Email, which provides social context to messages using a social network's underlying social graph. This not only gives each email recipient control over who can message him/her, but also provides the recipient with an understanding of where the message socially originated from. However, if a spammer compromises a legitimate user’s computer, the spammer can easily attack the user’s friends in the social network, which is characterized by high clustering and short paths [1].
E. Social network aided content and identity based spam filter (SOAP)

SOAP is a social network based personalized, attack-resilient, and user-friendly Bayesian spam filter. SOAP is combination of content based and identity based spam filtering. It further leverages social information including personal (dis)interests and social relationships. SOAP encourages users to indicate their (dis)interests and social relationships with their email correspondents in order to receive less spam and lose less legitimate emails. Using this information, nodes form a distributed overlay by connecting to their friends; that is, nodes use social network links directly as overlay links[1].

SOAP integrates four new components into the Bayesian filter: (1) social closeness-based spam filtering, (2) social interest-based spam filtering, (3) adaptive trust management, and (4) friend notification. Based on the collected social information, SOAP infers node closeness and email preference for individuals. The Bayesian filter keeps a list of spam keywords and their corresponding weights showing the probability that the email containing the keyword is spam. Based on the three social based components, after parsing the keywords of an email, SOAP adjusts the weights of the keywords. Then, SOAP resorts to the Bayesian filter for spam evaluation. The weights are therefore adjusted based on the closeness between the receiver and the sender, the receiver’s (dis)interests, the receiver’s trust of the sender, and the received spammer notification from friends. If the closeness is high, the likelihood that they send spam to each other is low, so the weight is decreased, and vice versa[1].

However, it is possible that close nodes are compromised. This problem is resolved by the trust management and friend notification components. For those nodes with low closeness, the emails are evaluated based on the user’s (dis)interests. The accurate results from SOAP become training data to automatically train the Bayesian filter, thus making the filter user-friendly and personalized, which also reduces the training time [1].

III. CONCLUSIONS

Internet Spam is unsolicited email messages sent over internet to annoy user. In this paper different techniques have given for spam filtering. Although no single technology can achieve one hundred percent spam detection with zero false positives, Machine learning , Neural Network , Bayesian filterinh , SOAP in particular have proven extremely effective and reliable at accurately identifying spam and minimizing errors to an acceptable minimum.

REFERENCES

[1] Haiying Shen, Senior Member, IEEE, and Ze Li, Student Member, IEEE, “Leveraging Social Networks for Effective Spam Filtering”, IEEE TRANSACTIONS ON COMPUTERS, VOL. 63, NO. 11, pp. 2743-2759, NOVEMBER 2014.

[2] GFI Software, “Why Bayesian Filtering Is the Most Effective Antispam Technology,” http://www.gfi.com/whitepapers/whybayesian-filtering.pdf, 2011.

[3] M. Uemura and T. Tabata, “Design, and Evaluation of a Bayesian-Filter-Based Image Spam Filtering Method,” Proc. Int’l Conf. Information Security and Assurance (ISA), pp. 46–51, 2008.

[4] J.S. Kong, P.O. Boykin, B.A. Rezaei, N. Sarshar, and V.P. Roychowdhury, “Let Your CyberAlter Ego Share Information and Manage Spam,” Univ. of California, Los Angeles, CA, technicalreport, 2005.

[5] H. Lam and D. Yeung, “A Learning Approach to Spam Detection Based on Social Networks,” Proc. Conf. Email and Anti-Spam (CEAS), pp. 1–10, 2007.

[6] https://en.wikipedia.org/wiki/Email_spam

[7] Amit Kumar Sharma, Sudesh Kumar Prajapati, Mohammed Aslem, “A Comparative Study between Naive Bayes and Neural Network (MLP) Classifier for Spam Email Detection”, IJCA (0975 – 8887) National Seminar on Recent Advances in Wireless Networks and Communications, NWNCC, pp.12-16,2014

[8] Peiman Mamani Barnaghi, Vahid Alizadeh Sahzabi and Azuraliza Abu Bakar,” A Comparative Study for Various Methods of Classification”, (ICICN) IPCSIT ,vol. 27,pp.62-66,2012.