Non Machine and Machine Learning Spam Filtering Techniques

Sandhi Kranthi Reddy, T Maruthi Padmaja

ABSTRACT--- Email is an effective communication method used in most of the organizations which is abused by spam. Spam email is an unwanted mail which leads to phishing websites. On an average a user on internet may get 10-15 spam emails per day. There are many effects of spam emails such as fills up user's inbox, consumes resources such as disk space and bandwidth, etc., may also contain attachments which corrupts users data. It is difficult to user to always check and decide whether the email is spam or not. Spam filtering mechanisms are used to detect spam emails. In this paper a detailed review is given how machine and non-machine learning techniques are used to detect spam emails.

Keywords: Spam, Ham, Spam Filtering Mechanism.

I. INTRODUCTION

As the rapid development in internet everyone is communicating through emails. With these usage of emails lot of spam mails are increasing day by day. Spam email is unwanted mail which contains information about particular product or any other services provided by the companies. On average a user on internet may get 10-15 spam emails per day [1]. There are lot of effects with spam emails such as consumes user's valuable time to read complete mail, fills up user's inbox, more than 70% of global email traffic consists of spam, consumes resources such as disk space and bandwidth, steals user information such as personal details, credential details etc. spam emails may also contains compressed attachments on extraction it corrupts complete user's data on machine [1]. To identity spam emails there are two methods namely, Non Machine Methods and Machine Learning methods.

II. NON MACHINE LEARNING SPAM FILTERS

Non machine learning methods are used to detect whether incoming mails are spam or not, these methods mainly includes list of email ids and list of words based on that it detects whether the given mail is spam or not[2][3]. The Non Machine methods includes list based spam filters and content based filters. List based filters identifies based on predefined list whereas content based identifies based on the content of the mail.

2.1 List Based Filter

List based filters includes black list, white list and grey list.

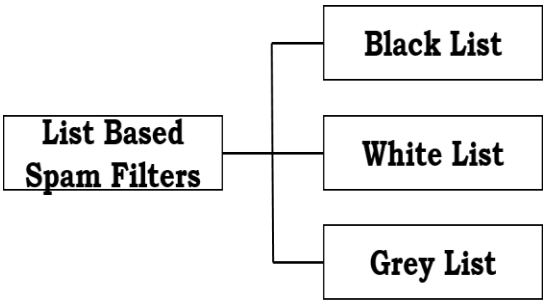

Figure 1 : Types of Spam Filters

2.1.1 Black List

Black List is the most common spam filtering technique which contains a predefined set of email addresses or IP addresses of known spammers(who have sent spam emails previously), So that the email from those addresses can be blocked. Whenever an email arrives the filter compares its email address or IP address with black list. If it present in the black list, then the email gets rejected to enter into inbox by considering it as spam. Black list can generate false positives if the spammer uses the IP address which was used by legitimate senders[2][3]. Even they can switch to different IP addresses or email address as they are also getting updated with the latest technologies.

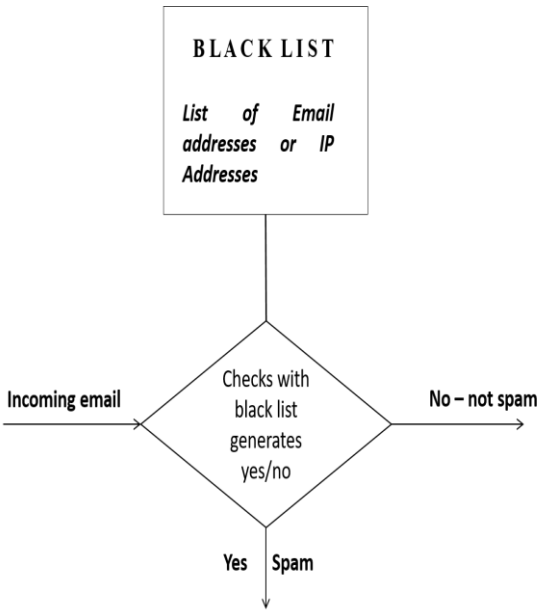

Figure 2 : Black List Spam Filter

Revised Manuscript Received on February 11 , 2019.

S. Karunya Research Scholar, Department of Computer Science & Engineering, Vignan's (Deemed to be University), Guntur, A.P, India. (E-mail: kranthi.sandhi@gmail.com)

Dr. T Maruthi Padmaja Associate Professor ,Department of Computer Science & Engineering, Vignan's (Deemed to be University), Guntur, , A.P, India. (E-mail: padmaja.tu2002@gmail.com)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

2.1.2 Real Time Black Hole List

Real time black hole list is identical to black list except that predefined list of spammers is maintained by the third party. Here the list is referred as real time black hole list. In this technique, simply filter should get connected with the third party system so that whenever an email arrives, the senders email or IP address gets compared with black hole list. Real-time black hole list also results in false positives similar to that of black list. In black list we have very less control over black hole list since it is maintained by third party.

2.1.3 White List

White list maintains a trusted-user list (which contains the email or IP address from where the user wants to get emails in inbox). It is advised to use a white list with another spam-filtering technique to avoid false positives because if we use only white list technique, any legitimate who are added to white list may get blocked automatically. One variation of these technique is automatic white list.[2][In this technique if an email from unknown sender arrives, its address is compared with databased if it has no spam history, the email is sent to inbox and address is added to white list.

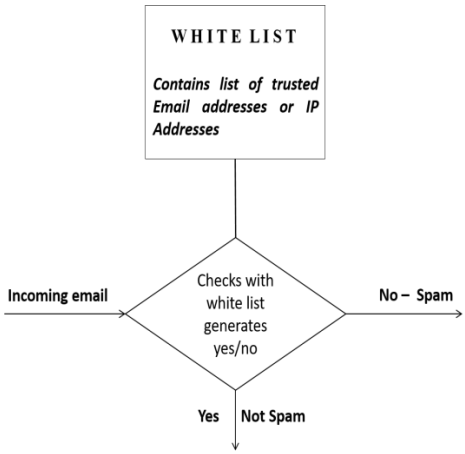

Figure 3 : White List Spam Filtering Technique

2.1.4 Grey List:

Grey list spam filtering technique will reject the emails from unknown senders and mails a failure message to the sender for the first time. If the sender resend the message for second time (most of the legitimate senders will do this) then the mail is sent to inbox and address is added to grey list. This is done since most of the spammers sends junk mails only once[3].

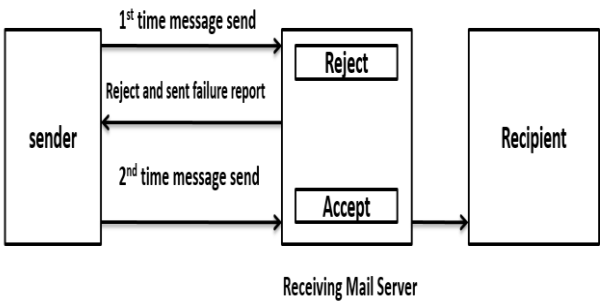

Figure 4 : Grey List Spam Filter

All the list based filters detects the spam mails based on the email id or IP address and delays mail delivery which creates inconvenience in case of receiving time-sensitive messages.

2.2 Content Based Filter

Content Based filters includes word-based filter and heuristic filter.

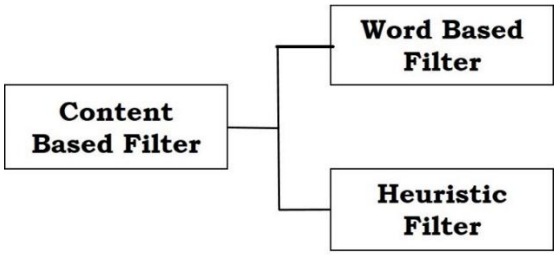

Figure 5: Content Based Filter

2.2.1 Word Based Filter

Word-based spam filter method mainly concentrates on the content of the email. It contains the list of words which are mostly found in spam emails. So whenever an email arrives it compares its content with the word list. If any word of mail matches list then mail is considered as spam. Word based filters also generate false positives. Word list contains commonly used words such as discount, cashback etc then mail may not receive from legitimate sender who are offering a product with some discount at reduced price. Spammers can also misspell words. List should be routinely updated, so it consumes time.

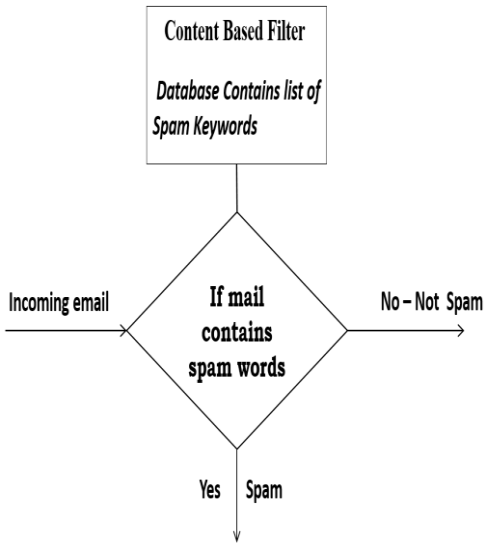

Figure 6 : Word Based Filter

2.2.2 Heuristic Filter

Heuristic filter is rule-based filters contains word list but here each word is assigned with priority. The words which are mostly used in spam messages are assigned with higher priority and the words mostly used in normal mails are assigned with lower priority. The filter should sum up all

priority values of words in mail to get total score. If total score is greater than the critical value then that mail is considered as spam else it is sent to inbox [3]. Heuristic filters work fast without any delay.

III. RESULTS & DISCUSSIONS

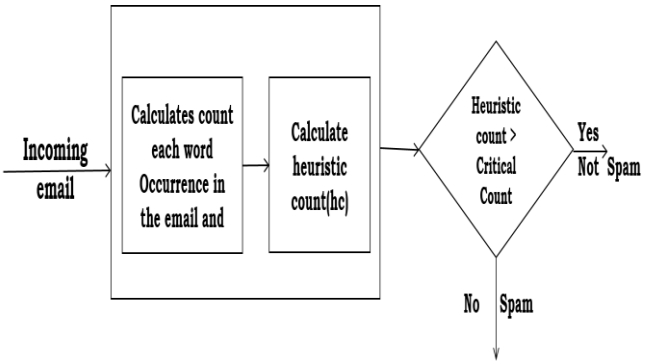

Figure 7: Heuristic Spam Filter

IV. MACHINE LEARNING SPAM FILTERS

Machine learning is one major branch in AI and it is used in almost all areas such as banking, e-commerce sites, recommendation systems, social networking sites etc. Machine learning makes the machine to learn and react accordingly when new input comes. To identify whether an arrived mail is spam or not many machine learning algorithms are used. The machine learning based spam filters mainly concentrates on header of the email or content of the email whereas list base filter checks either email id or ip address and draws conclusion whether the mail is spam or not these requires less computation but frequently list must be updated.

Machine Learning based filters can be considered as second category spam filter. collect existing data, denoted as "training data", and generates the models which is used to predict the newly arrived data.

- Some of the spam filtering methods are designed to detect spams by scanning fully the content of the email.
- Some of the spam filtering methods are designed to detect spams by scanning only header of the email.

3.1 Spam Filtering Method Detects the Spam Emails by Scanning Fully Content of the Email

- Bayesian filter which scans fully content of the mail.
- In 1996, Cohen Proposed a filtering method based on RIPPER learning algorithm**.**
- In 1999, Drucker et al. applied support vector machines
- In 2001 carreras and marquez introduced a new method based on AdaBoost algorithm
- In 2005, Delany and Cunningham presented a kind of KNN method.
- In 2007, He and Bo filtered spams by constructing a new asymmetric boosting method.
- In 2008, Hsiao and Chang constructed an incremental clustering-based filtering technique.

3.1.1 Bayesian Filters

Bayesian Filter is considered as the most advanced form of content-based filter. It employs the laws of mathematical probability to determines which are the legitimate emails and which are spam emails. Initially user must train the filter manually by flagging each email as either spam or legitimate. Over time, the filter takes words or phrases found in legitimate emails and adds them to a list; it does the same with the word or phrases found in spam emails. Whenever an email arrives the filter scans the contents of email and compares it against its two word-lists to calculate the probability to find whether it is spam or legitimate email[4].

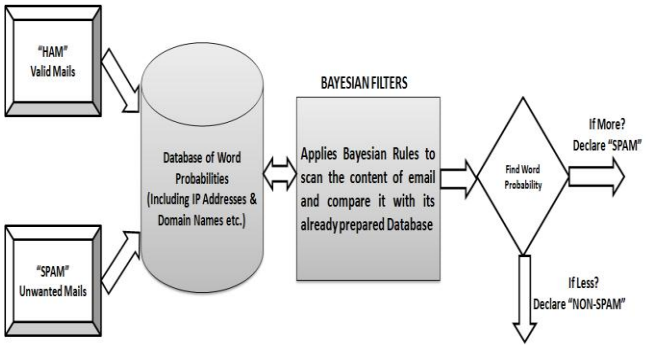

Figure 8: Bayesian Filter

For example, if the word 'claim' has appeared 62 times in spam messages list but only 3 times in legitimate emails, there is a 95% chance to consider the email containing the word 'claim' as spam. As a Bayesian filter constantly build its word list based on emails that a user receives, it theoretically becomes more effective the longer it's used. Since it does not require a training period before it starts working well, user should be with patience to manually delete spam emails at least at first.

Apart from Bayesian they are many machine learning algorithms which are applied on different data sets, In 1996 Cohen Proposed a filtering method based on RIPPER learning algorithm to classify the text and compared Ripper with TF_DF method. In 1999, Drucker et al. applied support vector machines to classify email[9] and also compared it with other three algorithms Ripper, Racchio and Boosting decision trees. Among four SVM given acceptable performance [6]. In 2001 carreras and marquez introduced a new method based on AdaBoost algorithm in which high level of F1 measures achieved and increasing the complexity of base learners allows to obtain better "high precision" classifier[7]. In 2007, He and Bo filtered spams by constructing a new asymmetric boosting method, in this method boosting done with different costs and applied it on email spam for each round of boosting it improved false negative rates at the low false positive region in the ROC curve. In 2008, Hsiao and Chang constructed an incremental clustering-based filtering technique, which contains two phases. In first phase, it clusters emails in each given class into several groups and in second phase we capacitate ICBC with an incremental learning mechanism that can identify spam emails. In 2014 Zhou et al prosed a cost-sensitive three-way spam filtering method which detects the email based on the cost-sensitive topic. The problem with above methods is they will detect whether the email is spam or not based on the content of the email.

3.2 Machine Learning Spam Filter Method are Designed to Detects the Spam Emails by Scanning Header of the Email

- ✓ In 2007, wang and chen proposed a method by using header session. [10]
- ✓ In 2009 sheu and chu developed an efficient filtering method which analyzed only email's header section[8].

Wang and chen proposed a statistical method on the content of 10,024 junk emails collected from a spam Achieve database, and 599 regular emails in company with 635 solicited or commercial emails. Content analysis results demonstrate that to 92.5% of junk emails are filtered out when utilizing the message-ID, Mail user agent, and sender and receiver address in the header session as cues.[6]

Sheu and chu developed an efficient filtering method which analysed only email's header section and identifies whether it is spam or not. This method had done in-depth analysis on Chinese spam and got 96.17% accuracy and precision of was upto 98%. [7]

All above spam filters are applied on different data sets i.e., Bayesian can be applied on based on bag of words, Chinese spam, public corpus, TREC data set etc. Each data set may contain different words and spam filtering methods accuracy is varying from one data set to another data set.

V. PROBLEMS IN STATIC METHODS

Most of spam filters of machine learning are designed to intercept spams in static environment. But the internet is a dynamic database and will generate constantly the problem of concept drift. The concept of data changes along with time. The continuous change of data along with time would result in the renewal of conceptual model. The current machine learning spam filters have some short coming in dealing with concept drift.

In 2017 Jyh-Jian Sheu and Ko-Tsung Chu developed An efficient Incremental Learning mechanism for tracking concept drift in spam filtering[1]. In this they addressed the above issues by designing a spam filtering mechanism based on machine learning method, in order to solve the concept drift. In this they applied decision tree data mining algorithm to learn rules about spams from training mails.

VI. CONCLUSION

To classify whether the email is spam or not we have two types of methods non-machine learning and machine learning.

Non-machine learning includes list based and content based. List based includes black list, real time black list white list grey list which blocks/allows the mail if sender email id or IP address exists/doesn't exist in the pre-existing list. In case of list based filters the concentration mainly on email id or IP address with these the list based filters may generates false positives and true negatives. Another type of non-machine learning approach is to compare the email content with set of pre-existing words if match is found with pre-existing word then the email is considered as spam otherwise it not spam. Non-Machine learning techniques are not effective because of email id or IP address and based on certain words it classify the email as spam or not spam. Sometimes it doesn't generates proper results and frequent updates are required for pre-existing list or list of words or word database.

Many Machine learning spam filtering are proposed to classify the spam emails by scanning either content or header session of the email. Till now all researchers are concentrated on how to block spam emails by scanning either email id/IP Address or header of the email or content of the email but apart from these they are many other spam emails containing only images, attachments without any header part or content of the email or even single word.

Now-a-days spammer are sending emails that contains only images without any header section. These type of emails are creating lots of problems to the users. Sometimes some unwanted images was sent from anonymous mail, some attachments which contains virus file which corrupts complete data in the computer and to retrieve back it may leads to provide credential information. Generally these type of problems occurs to organizational email ids i.e., web mail id's. The spam filters are required to design in way that along with images and attachments it has to scan the mail and decide whether it is spam or not.

VII. ACKNOWLEDGEMENT

We would like to show our gratitude to our university for supporting to do these work.

REFERENCES

1. Jyh-Jian Sheu1, KoTsung Chu2 An efficient incremental learning mechanism for tracking concept drift in spam filtering, PloS ONE 12(2): e0171518, dio:10, 1371/journal,pone.0171518.
2. Xin Liu, Pingjun, Zou, Weishan Zhang, Jiehan Zhou "Research Article CPSFS: A Credible Personalized Spam Filtering Scheme by Crowd sourcing", Hindawi, Wireless Communications and Mobile Computing Volume 2017, Article ID 1457870, 9 pages.
3. Saima Hasib, Mahak Motwani, Amit Saxena Anti-Spam Methodologies: A Comparative Study, IJCSIT, Vol. 3 (6), 2012,5341-5345
4. J. Jaeyeon and S Emil, "An Emprical Study of Spam Traffic and the use of DNS Black Lists", in Proceedings of the 4th ACM SIGCOMM Conference on Internet measurement, pp.370-375, October-2004.
5. William W. Cohen Learning Rules that Classify E-Mail, AAAI Technical Report SS-96-05.
6. Harris Drucker, Senior Member, IEEE, Donghui Wu,Student Member, IEEE, and Vladimir N. Vapnik Support Vector Machines for Spam Categorization, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 10, NO. 5, SEPTEMBER 1999
7. Xavier Carreras and Llui's Marque Boosting Trees for Anti-Spam Email Filtering, RANLP-2001, pp. 58-64, Bulgaria, 2001

Jyh-Jian Sheu, KoTsung Chu An Efficient Spam Filtering Method by Analyzing E-Mail's Header Session Only, International Journal of Innovative Computing, Information and Control ICIC International c 2009 ISSN 1349-4198 Volume 5, Number 11(A), November 2009.

8. Chih-Chin Lai An empirical study of three machine learning methods for spam filtering, Knowledge-Based Systems 20 (2007) 249–254, Elsevier.
9. Chih-ChienWang Sheng-YiChen "Using header session messages to anti-spamming" , Elsevier, Computers & Security, Volume 26, issue 5 August-2007, Pages: 381-390