RESEARCH ARTICLE OPEN ACCESS

A Review on Different Spam Detection Approaches

[1]K.Ravikumar, [2]P.Gandhimathi
[1]Asst.professor, Dept.of.Computer Science, Tamil University, Thanjavur-613010.
[2]Research Scholar, Dept.of.Computer Science, Tamil University, Thanjavur-613010.

----------**********----------

Abstract:

The spam mails are used by spammers to pilfer the data of the users and organizations online. Quick growth rate of the use of the internet has increased the spam mails. There are several methods employed for filtering spam. Most of the spam filtering techniques are based on objective methods such as the content filtering and DNS/reverse DNS checks. Recently, some cooperative subjective spam filtering techniques are proposed. Objective methods suffer from the false positive and false negative classification. Objective methods based on the content filtering are time con-suming and resource demanding. They are inaccurate and require continuous update to cope with newly invented spammer's tricks. On the other side, the existing subjective proposals have some drawbacks like the attacks from mali-cious users that make them unreliable and the privacy. In this paper, we propose an efficient spam filtering system that is based on a smart cooperative subjective technique for content filtering in addition to the fastest and the most reliable non-content-based objective methods. The growing use of internet has promoted an easy and fast way of e-communication. The well-known example for this is E-MAIL. The increasing use of e-mail and the growing trend of Internet users sending unsolicited bulk e-mail, the need for an antispam filtering or have created, Filter large poster have been produced in this area, each with its own method and some parameters are to recognize spam.

Keywords — **Spam Filtering, E-mail, Images, Text Tagging, Web Application Web Services.**

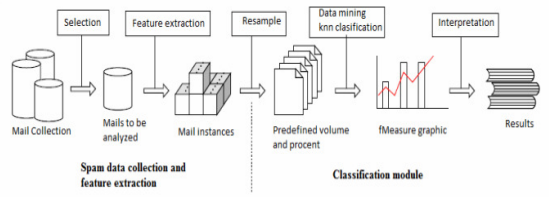
----------**********----------

I. INTRODUCTION

Email Communication is indispensable in the present days. Spam mails are unsolicited junk mails sent by the spammers which adversely affect the email communication process. Unsolicited commercial mails are often sent by the spammer to illegally promote a service or product. Spam became an issue when the internet was opened to the public. Internet users are forced to receive spam mails in their inbox. Most of the spam filtering techniques are based on objective methods such as the content filtering and DNS/reverse DNS checks. Recently, some cooperative subjective spam filtering techniques are proposed. Objective methods suffer from the false positive and false negative classification. Objective methods based on the content filtering are time con-suming and resource demanding. They are inaccurate and require continuous update to cope with newly invented spammer's tricks. Moreover, the association of spammers with hackers and virus writers poses a very real threat to the Internet security and availability. About 8 years ago, spam was sent by spammer's own e-mail servers. Approximately 45% - 60% of spam is now sent from compromised systems distributed over the Internet. Spam relaying increases the distribution base and at the same time eludes and overwhelms spam detection systems.

Many filtering techniques have been developed to control the flow of spam emails. Unfortunately, even with

these available techniques, the number of spam emails is growing and the flow has not been controlled completely. The setback is that there is no actual solution because a spammer; an unidentified user with enough knowledge is able to be familiar with the logic of the filtering mechanisms.



Spams are always changing their contents and forms, so that the anti-spams can't realize them. Some methods to prevent propagation of spams are including:

- Economic methods: pay to send emails: like email protocols legislative methods: such as can-spam law, secure email transfer bed.
- Change email transfer protocols and offer alternative protocols such as sending ID.
- Control output and input emails
- Filtering based on learning (statistics) by using mail features
- Detecting a phishing mail (fraud page) by the help of fuzzy classification methods.

II. LITERATURE SURVEY

The solutions using both machine and non –machine learning approaches are reviewed and taxonomy of different approaches is presented. While a range of different techniques have and continue to be evaluated in academic research, heuristic and Bayesian filtering, along with its variants provide the greatest potential for future spam prevention. M.Basavaraju and Dr. R. Prabhakar performed a work," A Novel Method of Spam Mail Detection using Text Based Clustering Approach". A new spam detection technique using the text clustering based on vector space model is proposed in this research paper. By using this method, one can extract spam/non-spam email and detect the spam email efficiently. Representation of data is done using a vector space model. An approach which is character-based technique. This approach uses a multi-neural networks classifier. Each neural network is trained based on a normalized weight obtained from the ASCII value of the word characters. Results of the experiment show high false positive and low true negative percentages. R. Kishore Kumar, G. Poonkuzhali, P. Sudhakar provides the analysis of email spam classifier through data mining techniques. In their work," Comparative Study on Email Spam Classifier using Data Mining Techniques " spam dataset is analyzed using TANAGRA data mining tool to explore the efficient classifier for email spam classification.

III. List Based or Rule Based Filters

List based filter attempt to stop spam by categorizing senders as spammers or trusted users, and blocking or allowing their messages accordingly.

3.1. Blacklist

Black list is the form of rule based filtering that uses one rule to decide which emails are spams. Black list are the list of IP address of machine or record of email addresses that have been previously used to send spam. When incoming message arrives, the spam filter checks to see if it's IP or email address is on the black list, if so, the message is considered spam and rejected. Blacklist can be used for on both large scale and small scales. Advantage is it can block substantial amount of email. Disadvantage is a blacklist provider can block an entire net block range instead of just an individual IP.

3.2.White list/Verification Filter

While Blacklisting is used to decide which emails are spam, but White listing is used to decide which emails are ham and assume all other emails are spam.

3.3. Blackholes

Spam Blackholes work hand in hand with Blacklist. The way Blackholes work is someone posts message on websites, Usenet, forum, etc, showing their email address. The email address they use is generally a machine account that detects who sent the spam and the IP address of to a DNS Blacklist.

IV. EFFICIENT SPAM DETECTION METHOD

The methods currently used by most anti-spam spam software are static, mean that it is fairly easy to evade by tweaking the message little. To do this spammer simply examines the latest anti spam techniques and find the ways how to dodge them. To effectively combat spam, an adaptive new technique is needed. This method must be familiar with spammer's tactics as they change over time. It must also able to adapt to the particular organization that it is protecting for the answer lies in Bayesian mathematics. Why Bayesian filtering is better. The Bayesian method takes the whole message into account- It recognizes key words that identify spam, but it also recognizes words that denotes valid mail. A Bayesian filter is constantly self adapting – By learning from new spam and valid outbound mails, the Bayesian filter evolves and adapts to new spam techniques. The Bayesian method is sensitive to the user. The Bayesian filter is multi-lingual and international- A Bayesian anti-spam filter, being adaptive, can be used for any language required.

V. CONCLUSION

We combine two methods to detected the mail is spam or legitimate mail. OCR algorithm is used for converting image (from emails) to text. Bayesian Algorithm is used to detect the probability that the words in email are spam or not. Based on the Bayesian algorithm email is added to spam list (without user interaction) and only the legitimate emails are shown to the user. This implies that if an email is identified as spam one, the receiver's bandwidth and memory is preserved which will assure a better performance. Finally, by locating the filtering system in the sender mail server; the processed time becomes *n* times less than the time when the filtering system is in the receiver mail server when *n* indicates the number of processed emails.

VI.REFERENCE

[1] Godwin Caruana, Maozhen Li, Yang Liu, "An ontology enhanced parallel SVM for scalable spam filter training," *Neurocomputing Elsevier*, vol. 108, pp. 45-57, 2013.

[2] E. Blanzieri, A. Bryl, "A survey of learning-based techniques of email spam filtering," *Artificial Intelligence Revolution,* vol. 29, pp.63–92, 2008.

[3] L. Zhang, J. Zhu, T. Yao, "An evaluation of statistical spam filtering techniques," ACM Transaction on Asian Language Information Process, vol. 3, pp.243–269, 2004.

[4] H. P. Graf, E. Cosatto, L. L. Bottou, I. Durdanovic, V.Vapnik, "Parallel support vector machines: the cascade SVM," *In Advancement of Neural Information Process System,* pp.521–528, 2004.

[5] J. Dean, S. Ghemawat, "MapReduce," *Communication of ACM*, vol. 51, pp.107, 2008.

[6] B.He, W.Fang, Q.Luo, N.K.Govindaraju, and T.Wang, "Mars: a MapReduce framework on graphics processors," *Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques(PACT),* pp.260–269, 2008.

[7] Naïve Bayes classifier. [Online]. Available:http://en.wikipedia.org/wiki/Naive_Bayes_classifier.

[8] POPFile - Automatic Email Classification- Trac. [Online]. Available: http://www.getpopfile.org/.
[9] Uber spamfence: spamfence powered by eleven.[Online]. Available:http://www.spamfence.net/about spamfence.
html.
[10] Spamihilator – Free Anti-Spam-Filter. [Online]. Available: http://www.spamihilator.com/.