IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A CLIENT BASED ANTI-SPAM FILTER

Minimol.R[1], Thahira Banu.V[2]

[1]Student, I M.Sc Computer Science, Sri Krishna Arts and Science College, Tamil Nadu

[2]Assistant professor in Computer Science, Sri Krishna Arts and Science College, Tamil Nadu

ABSTRACT

The "Anti-Spam Filter" which helps in the process of detecting the spam in the emails. In an ordinary mail system, it contains only the default detection of the spam emails. In order to avoid limitations in the existing system, proposed system aims of creating a web application. In which the system can detect types of spam such as Nigerian letter, advertising, phishing, and chain letters etc. At first, the system analyses the listing methodologies such as standard, primary, important. After that, it verifies whether it is spam or not. If its spam, again it checks with the types of category. So that it moved into the spam list and shown in a hidden format. Otherwise, it will move to the inbox emails. Finally, this will remove all category of unwanted spam mail and shows the original emails.

Keywords: **spam; email; fuzzy classification; listing methodologies in spam;**

I. INTRODUCTION

Email is an active and competitive way of communication. The fast development of information and communication technology brings not only advantages but also certain issues, for example, spam. Spammed email, also known as junk email, it is uninvited messages sent in bulk through email. The name comes from ***Spam luncheon*** meant by way of a *Monty Python* sketch in which Spam is everywhere, unavoidable and repetitive.

Email spam has whether it is commercial or not, they are not only disturbing but also dangerous because they may contain links that lead to phishing web sites or sites that are hosting malicious or include malicious as file attachments. Depending on the aim and aspiration of the sender, spam messages may or may not include commercial information. That is why two groups of messages, such as uninvited commercial emails and unsolicited bulk emails declared. Backscatter spam, Image spam, Botnet spam are few types of spam[5]. In text spam filters, the words such as "Special offers", "discount", "You have won in lottery" are the excellent signal of spam messages. Spam becomes a nightmare for Internet users [4]. This paper presents the process of generating a client-based anti-spam filter program which justifies the question like: "How to avoid spam without killing an actual electronic message?". Using of listing techniques it will reduce the time but increases the accuracy of identifying spam messages. Listing techniques such as

1. Black listing
2. White listing
3. Gray listing

1. **Black listing**: A database is maintained by either client or server to store the details about known and unknown abusers, addresses
2. **White listing**: The mailer program allows mails only from the contact users. If any unknown users want to send messages, they should request first to send the mails. This method blocks unknown email addressed messages.
3. **Gray listing**: Whenever unknown messages are received it blocks the message and also generates and reports error message to that server.

Spam gives essential background information that used to create the project, such as the definition of spam, different kinds of spam, techniques to avoid spam messages and tools (software programs) that use given techniques. Determining spam categories improve the capability of spam filtering techniques especially of the context filter.

THE IMPACT OF SPAM

Mass mailing spam has a low cost. However, a serious amount of useless messages cause visible harm to recipients. First of all, it is all about time wasting on needless screening and sorting important messages. Also, the Internet service is expensive, and the user has to pay for unwanted messages. It seems that spam can be favorable for providers, as it increases the amount of traffic. But, in fact, providers also incur an additional cost due to the increased capacity of useless channels and tools**[1]**.

CATEGORIES OF SPAM

1. **ADVERTISING SPAM**
The advertising message always acclaims real goods or services and provides links to the original location, where users can get more information about them. Site promotion spam advertises remarkably good and free products by providing a link to the site that does not have any reasonable information. Advertising spam used to charge calls announce a product and provides a telephone number so that if users call that number, they will receive a bill. Marketing probe spam usually asks users to complete a questionnaire and send the information to a specified address. However, some advertising spam includes viruses that affect a user's system. Usually, spam advertises a product or services and marks up the counter on the website.

2. **NIGERIAN LETTER**
Sometimes spammers use spam to lure fund from the message recipient. The most familiar method is called the "Nigerian letter". Nigerian letters contain a text which lectures about money that the recipient can get from sender under certain circumstances: for example, for helping the sender to send some money to organize documents. Luring money from victims is the purpose of this type of fraud.

3. **PHISHING**
It is a scam using spam messages. Phishing is an attack to defraud the recipient by taking decisive information details such as credit card numbers, passwords or accessing email accounts. This kind of message usually disguised as a formal letter from the administration and asks the recipient to provide information by following the link provided by the spammer. Often phishers use emotive writing form or pretend to be licensed party to gain the trust of their reality to the receiver. Sometimes phishing spam includes get-rich schemes.

4. **CHAIN LETTERS**
It is another spam that encourages recreating letters and sending it to many beneficiaries as the user can find. In this, the receiver with consequences of the conditions in which the message are not met, or inspires by benefits that the user will get if the conditions are met. Usually, the letter starts with the well-defined phrase as "It is important! Please, do not remove, read it". Chain letters do not cause a financial loss like a fraudulent pyramid scheme, However, they do fill the mailbox with useless messages and make the user pay for the traffic generated by spam.

II. LITERATURE REVIEW

In order to start building the entire project, all information related to spam messages and influences to the people and techniques to overcome problems which are caused by spam was found then proper information is collected by reading and analyzing scientific papers and books.

An article about spam filtering method and their improvements was written by final year students "A Content-Based Classification of Spam Mails with Fuzzy Word Ranking", the authors G.Santhi[1], S. Maria Wenisch, and Dr. P. Sengutuvan were discussing to find the spammed content and their ranking based on the word which is predefined in that project. Another research paper was written by Nazerke Rakhymbayeva a final year student. In this article
"A client-based anti-spam filter" has shown the different methods to avoid spam content. These two papers have not only advantages, but it also has some drawbacks,

- It uses more than two techniques and methodologies to find the spam content.
- It uses the same techniques for both grown since the early 1990s, and by 2014 was predicted that it made up around 90% of email messages sent. Since the liability of the spam is borne mostly by the beneficiary, it is effectively postage due to advertising. This makes it an excellent example of a negative expanse. Most email spam messages are commercial in nature.

EXISTING SYSTEM

In the existing system, while sending information from one user to another user through the mail, the spammed mail is the unwanted content which cannot be needed for the user. So, that it can be categorized into the spam category. So, that user can view in case of information needed by the user.

LIMITATIONS

- It can avoid only some kind of spam data.
The user can view any spammed contents.
- It is a loss of time
- It has used the ranking system based on words but it does not work effectively all the time.

Table : Review of Spam filtering Techniques

AUTHOR	TITTLE	CONTENT AND REF NO	TECNIQUES USED
1. Nazerke akhymbayeva	“A client-based anti-spam filter”	Categories of spam , spam methods, detection of Spam and listing methodologies[1].	1. Listing techniques.
1. G.Santhi 2. S. Maria Wenisch 3.Dr.P. Sengutuvan	“ A Content Based Classification of Spam Mails with Word Fuzzy Ranking"	The fuzzy Classification Contains fuzzy Inference system and a set of rules. The Proposed fuzzy rule Based classification. This work has a list spam words in the database with its ranked value [2].	1.fuzzy classification Module 2.Ranking of words 3.classification of spam mails.
1.Riya Mehta 2.Ankita Gandhi	“SMS Spam Filtering”	The Definition of Spam SMS does not very much in Case of Emails or SMS Spam in simple word the it can be Describe as “Unsolicited Bulk Message” these are unwanted for the user Sent by Sender due to low price the company and Spammer used this service for marketing and promotion[3].	1. filtering techniques 2. Text classification

PROPOSED SYSTEM

The proposed system is designed for avoiding all the categories of the spammed emails. There different types of categories such as Nigerian letters, advertising spam, phishing email, and chain letters. The spammed mail can be hidden from the user. In case they want to see those spam emails at that time they can view their details.

ADVANTAGES

- The spam contents can be hidden.
- Unrelated information can be shown to the user.

LISTING METHODOLOGIES

It is used to categorize the mail in Inbox. It consists of three methods. They are:

- **Blacklist** - It rejects all emails from the sender that is predefined in this list.
- **Whitelist** - It allows all incoming emails from the sender that is predefined in this list.
- **Graylist** - It will temporarily reject the incoming emails from unknown sources and checks whether it is spam or not using fuzzy classification[1].

III. SPAM DETECTING METHOD

It consists of three steps they are :

(i)First, the mail is categorized by three types such as primary, social, updates and store in the inbox by using the listing methodologies.

(ii)Then it checks for spam by using fuzzy classification.

(iii)If it spam content then it will be stored in spam folder otherwise it will be stored in the inbox.

FUZZY CLASSIFICATION

This arrangement is the procedure of grouping the item into a fuzzy data file whose membership function is calculated by the accuracy value of a fuzzy propositional function. A fuzzy propositional function is, linked to a statement involving one or more variables, such that, when values are hired to these variables, the expression becomes a fuzzy proposition.

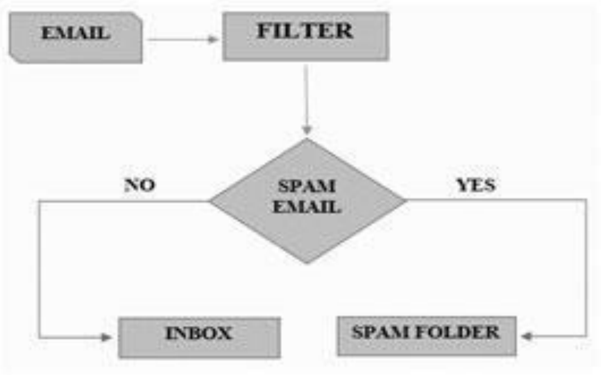
Accordingly, fuzzy classification is the technique of grouping individuals having the same characteristics into a *fuzzy set*.

A fuzzy classification compares to a membership function μ that illustrates whether an individual is a member of a class**[2].**

$$\mu \ :\mathbf{PF}\times\mathbf{U}\longrightarrow\mathbf{T}$$

μ – membership function

PF - accuracy value of the propositional function
x – condition operator
U- union
T-output in true value



Working of fuzzy classification

The fuzzy classification contains data set which is declared by keyword of variables. This classification contains a list of spam words stored in the database with predefined keywords. The spam words are derived from the content of the email. The spam words are assigned a value and categorized into four linguistic variables i.e. advertisement (ad), Nigerian letter (nl), phishing (ps), and chain letter (cl). This work obtains spam words from emails such as winner, dollar, award, cash prize, top job opportunities, earn more, beneficiary, good news, claims, high salary, and payment is the few most interesting words used by the spammers to cheat the users. The internet user gets attracted by these words in the mail and connects the sender immediately and gets deceived. So these words are evaluated as very strong spam words. The actual content is extracted from the user inbox and compared against the spam data list in the database, If the actual content is equal to the spam content then corresponding mail is stored in. In order to classify the spam fuzzy in the spam folder in encrypted form**[3].**

Algorithm: E-mail Fuzzy classifier

Input: Email Testing Dataset

Step1: Start

Step2: Assign the input as keywords for

I. ad – Advertisement
II. nl-Nigerian letter
III. ps-Phishn
IV. cl-Chain letter

Step3: Select the mail using listing methodologies it categorizes and stores it in inbox database

Step4: Compare the mail with spam content

Step5: Plot the variables in assigned input keywords

Step6: Verify the condition [if(con==spam_con)]it checks for the spam, if the condition is true it will be encrypted and store in the spam mail database

Step7: Otherwise, it will be stored in the inbox database

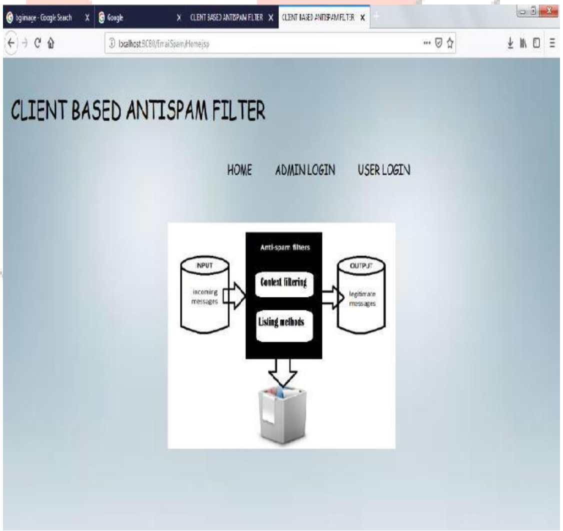
Step8: Display the result.

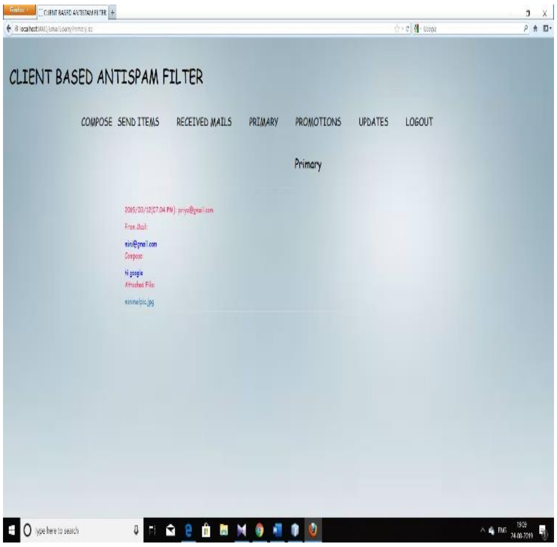
RESULT

This approach helps the end-users to identify the spam emails by using the spam datasets i.e., advertisement, Nigerian letter, a chain letter. The user can easily distinguish the spam emails and delete the spam emails in the inbox level and store in the spam folder in the hidden format.

CONCLUSION

This paper is a content-based classification of spam emails using fuzzy classification. There are many classifiers and filters available for arranging and cleaning spam emails. This paper analyses the previous related works. This filter method extracts only the character from the content of an email instead of extracting all the character from the mail. This approach helps the end-users to identify the spam emails by using the predefined data sets
i.e., advertisement, Nigerian letters, phishing, and chain letters. The user can easily distinguish the spam emails and delete the spam emails in the inbox level.





CLIENT BASED ANTISPAM FILTER

Primary

2019/03/12(07.04 PM): priya@gmail.com

From Mail:

nini@gmail.com

Compose:

hi google

Attached File:

nininelpic.jpg

FUTURE SCOPE

The actual content is extracted from the inbox of an email are compared with a list of inbox mail data sets in the database and the content are categorized according to its spam content. Fuzzy classification method classifies the spam and produces the output. This work obtains a better result for classifying the spam mail or not. The future work aims at the classification of spam content in the subject and HTML also.

REFERENCE

1. Nazerke Rakhymbayeva , BSc (Hons) Computer Science , Final year project , “A client-based anti-spam filter” [online].

2. G.Santhi1, S. Maria Wenisch and Dr. P. Sengutuvan, CRD, Prist University, Thanjavur, India, Department of Information Scienceand Technology, Anna University, Chennai, India,

3. “A Content Based Classification of Spam Mails with Fuzzy Word Ranking”[oneline].
4. Riya Mhta, Ankita Gandhi, Department of Computer cience & Engineering Parul Institute of Engineerig & Technology, Vadodara, Gujarat, India, ”ASurvey SMS Spam filtering”[online]
5. Saadat Nazirova, “Survey on Spam Filtering Techniques”, Communications and Network, 2011, 3, 153-160 doi:10.4236/cn.2011.33019 Published Online August 2011 (http://www.SciRP.org/journal/cn)