Enhancing Secure Email Communication using Tracking and Blocking of Suspicious Email User

S. Srigayathri
UG Student
Department of Computer Science & Engineering
Valliammai Engineering College

V. Udhaya Lakshmi
UG Student
Department of Computer Science & Engineering
Valliammai Engineering College

G. Rathiya
UG Student
Department of Computer Science & Engineering
Valliammai Engineering College

G. Sangeetha
Assistant Professor
Department of Computer Science & Engineering
Valliammai Engineering College

Abstract

E-mail spam continues to become a problem on the Internet. Spammed e-mail may contain many copies of the same message, commercial advertisement or other irrelevant posts. In previous system, different filtering techniques are used to detect these e-mails using random forest and neural network. Initially the detected spam's are directed to the spam folder and later it can be blocked by the mail user. This paper proposes to filter out the spam messages efficiently using SWEET protocol, which is a highly available censorship resistant infrastructure. Whenever, the client composes the mail, it is sent to the sweet server through email agent. At the sweet server, based on the message similarity using Naive Bayes algorithm, bunch of unsolicited bulk email could be filtered out on the server and blocked before it reaches the Gmail server. The IP address of that spam mail is tracked and blocked from further reception of spam mail.

Keywords: Sweet Protocol, Unsolicited Mail, Message Similarity, Naive Bayes

I. INTRODUCTION

E-mail is defined as the transmission of messages over communication networks E-mail is it is fast, flexible, and reliable. Computer-based mail and messaging became possible with the advent of time-sharing computers in the early 1960s, and informal methods of using shared files to pass messages were soon expanded into the first mail systems. The standard E-mail protocols are SMTP, POP3 and IMAP. The main issue in E-mail is spam mails. Email spam usually pertains to unsolicited commercial messages sent in bulk by people you don't know–although there are exceptions to this rule. Some spammers will argue that email spam is not any different than traditional junk mail, but there is one undeniable difference which is cost. The lack of significant barriers (cost) to entry (sending) is often cited as a key problem with email spam. Email filtering is the processing of email to organize it according to specified criteria. Most often this refers to the automatic processing of incoming messages, but the term also applies to the intervention of human intelligence in addition to anti-spam techniques, and to outgoing emails as well as those being received.

II. LITERATURE SURVEY

S.NO	*TITLE*	*AUTHOR*	*CONTENT*	*ADVANTAGE*	*DISADVANTAGES*
1	*Feature Selection and Similarity Coefficient Based Method for Email Spam Filtering*	*Ali Ahmed A.Abdelrahim, Ammar Ahmed E. Elhadi, Hamza Ibrahim, 1Naser Elmisba*	*Statistical feature selection approach combined with similarity coefficients are used to improve the accuracy and detection rate for the spam detection and filtering.*	*It enhanced the detection rate, false alarm rate and the accuracy. Degree of similarity between spam to spam samples was increased.*	*Sometimes ham mails are identified has spam due to preprocessing. Working only based on predefined samples.*
2	*Spam Filtering Email Classification (SFECM) using Gain and Graph Mining Algorithm*	*M.K. Chae , Abeer Alsadoon, P.W.C. Prasad, A. Elchouemi*	*a hybrid solution of spam email classifier using context based classification model as main algorithm complimented by information gain calculation to increase accuracy*	*reduce the processing time of the spam classification because processing time of 0.1 second and accuracy as improved*	*Not all the emails are classified using graphs. manual processing required*

3	*Efficient Spam Email Filtering using Adaptive Ontology*	*Seongwook Youn and Dennis McLeod*	*Ontology's allow for machine-understandable semantics of data. Resource Description Framework (RDF) which would be the form of "Subject – Object – Predicate"*	*The accuracy of the decision tree was approximately 97.17%, better classification algorithm*	*Producing result using decision tree was difficult. It doesn't produce results for unknown sets*

III. PROPOSED SYSTEM

In proposed system, the spam emails are blocked even before it reaches the mail server.

Initially the client have to login to the local host server, when the user composes the mail, then it checks for probability condition of naïve bayes and if the probability exceeds greater than 0.5 then automatically blocks the mail and if it is less than 0.5 it sends the mail to corresponding recipient. There will be also a feature called grouping of mails under a single group name. When the user wants to send the mail to group of users, he will simply type group name in the recipient address. All the group users will be acknowledged the he/she was added to the group. Once the mail is blocked, it automatically tracks IP address of the spammer and blocks that IP address. By using our mail system it will reduces the over utilization of system resources, reduce the storage usage, also reduces the network traffic

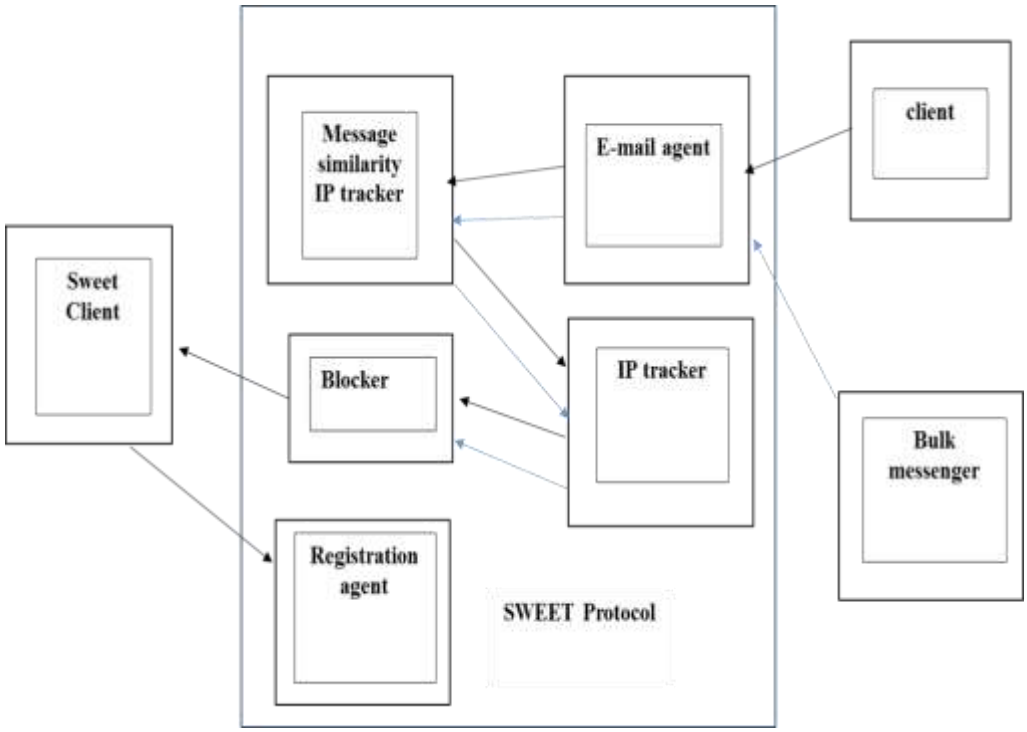


Fig. 1: Architectural Diagram

A. SMTP Protocol

SMTP (Simple Mail Transfer Protocol) is a TCP/IP protocol used in sending and receiving e-mail. SMTP usually is implemented to operate over Internet port 25. Many mail servers now support Extended Simple Mail Transfer Protocol (ESMTP), which allows multimedia files to be delivered as e-mail. You can use the IIS SMTP mail relay service to prevent spammers from directly interacting with your Microsoft Exchange Server! Internet users can send messages directly to your Exchange Server

B. IMAP Server

IMAP (Internet Message Access Protocol) is a standard email protocol that stores email messages on a mail server, but allows the end user to view and manipulate the messages as though they were stored locally on the end user's computing device(s). This allows users to organize messages into folders, have multiple client applications know which messages have been read, flag messages for urgency or follow-up and save draft messages on the server.

C. POP3 Protocol

Post Office Protocol version 3 (POP3) is a standard mail protocol used to receive emails from a remote server to a local email client. POP3 allows you to download email messages on your local computer and read them even when you are offline. Messages are stored on your local computer, which reduces the space your email account uses on your web server.

D. Naive Bayes Algorithm

Naive Bayes classifier is a straightforward and powerful algorithm for the **classification** task. Even if we are working on a data set with millions of records with some attributes, it is suggested to try Naive Bayes approach. Naive Bayes classifier gives great results when we use it for textual data analysis. Such as Natural Language Processing.
Below is the formula for calculating the conditional probability.
P(S/W)=P(W/S)*P(S)/ P(W/S)*P(S)+P(W/H)*P(H)
where:
P(S/W) is the probability that a message is a spam, knowing that the words occurs in it;
P(S) is the overall probability that any given message is spam;
P(W/S) is the probability that the word appears in spam messages;
P(H) is the overall probability that any given message is not spam (is "ham");
P(W/H) is the probability that the word appears in ham messages.

1) Advantages of Proposed Work

It reduces the traffic in the network.
Unwanted storage occupation by spam mails is reduced.
Blocking spam also keeps networks and servers running smoothly and efficiently. It prevents the systems from viruses and malwares attack by blocking spam mails.

IV. RESULT

Experiment results on a different spam filtering study shows that the many of the filtering technique is based on blacklist, white list and random forest out of which none produces ideal solution. The proposed approach is based on naïve bayes algorithm which leads to accurate identification of spam messages and is reliable. And also we provided a technique which could detect the exact spam messages and tracks the location of the spammer. Finally blocks the ip address of the spammer. Email spam filters using this approach can reduce the amount of spam messages and also can reduce the risk ofproductivity loss, bandwidth and storage usage.

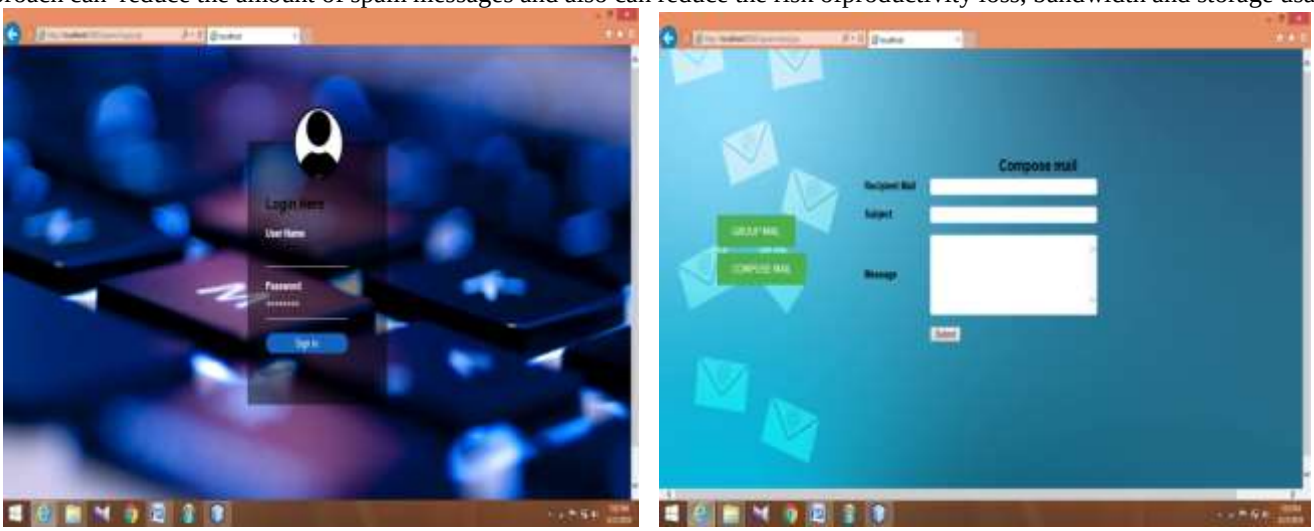
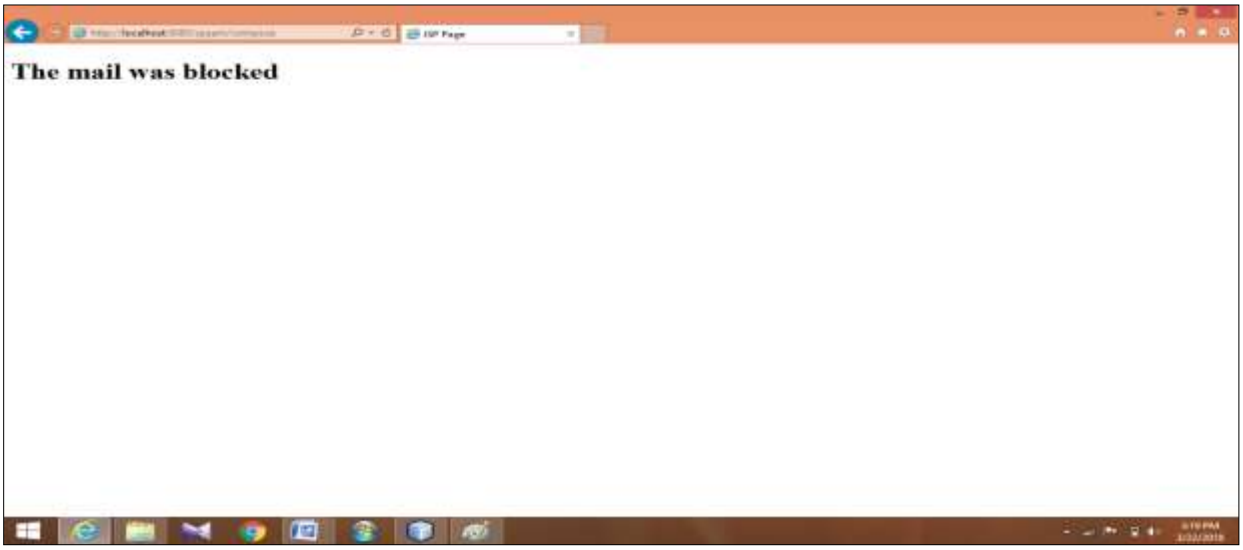


Fig. 2: Screen shots of result

```
Probality value:0.07692307692307693Count25
Probality value:0.038461538461538464Count25Ham count1
Probality value:0.6666666666666666
select email,epass from temp
Emailtestspam703passwordspam1234
hi user, we are selected for this offer.Due to this you will be provided with 50% discount.
not an group mail
IP address of Localhost is 127.0.0.1
Host name of your machine is User
insert into ipmanag(ipadr,hostname) values('127.0.0.1','User')
ksfnesngons
```



V. CONCLUSION

We have presented a spam filtering technique to identify spam messages from the mail. Unsolicited junk mail or spam is a major problem to companies and mail users. The proposed method is a processing step to detect spam messages. The accurate identification of spam messages is a key to block spammers. The problem is modelled as finding the spam messages from the mail by comparing body of the mail with the datasets using naïve bayes algorithm. Both ham and spam messages are contained in the dataset which helps to detect spam messages accurately.

REFERENCE

[1] Ali Ahmed A.Abdelrahim, Ammar Ahmed E. Elhadi, Hamza Ibrahim, 1Naser Elmisbah "Feature Selection and Similarity Coefficient Based Method for Email Spam Filtering", 2013 International Conference on Computing, Electrical and Electronic Engineering (icceee).
[2] M. K. Chae 1 , Abeer Alsadoon1 , P.W.C. Prasad1 , A. Elchouemi2 1 Charles Sturt University Study Centre, Sydney, Australia 2 Walden University, USA," Spam Filtering Email Classification (SFECM) using Gain and Graph Mining Algorithm", 978-1-5090-4228-9/17/$31.00 ©2017 IEEE
[3] Seongwook Youn and Dennis McLeod," Efficient Spam Email Filtering using Adaptive Ontology", International Conference on Information Technology (ITNG'07)
[4] Reshma Varghese, Dhanya K.A "Efficient Feature Set for Spam Email Filtering", 2017 IEEE 7th International Advance Computing Conference
[5] Wu-ying liu , Lin wang , Ting wang," Online supervised learning from multi-field documents for email spam filtering", Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010