ISSN 2395-1621

Security Email Filtering Using User Prediction

[#1]Swapnil Mahale, [#2]Shyam Manzarte, [#3]Kedar Joshi, [#4]Shridhan Patil

[1]swpnl.mahale@gmail.com
[2]manzarteshyam@gmail.com
[3]kedar.joshi2012@rediffmail.com
[4]pshridhan@gmail.com

[#1234]Deparment of Computer Engineering
JSPM's
Imperial College Of Engineering & Research,
Wagholi, Pune – 412207

ABSTRACT

Modern Developers already having the problem of classifying the text and identifying the spam emails. Now days, spammers are intelligent enough to develop a new technologies and types of email content becomes more different, text based anti-spam approach alone are not enough in preventing spam. For endeavour to overcome the anti-spam development technologies hackers are recently using the image spam technologies to make the analysis of email's body text incompetent. The primary purpose of this project is to develop efficient and strong spam discovery system. The structure we have developed will be enable to identify and discover the data of emails, specially the image generated by spammers artificially sent as attachment in an email. The system will analyse the image content and classify the embedded image as spam or legitimate hence classify the email accordingly.

Keyword: **Spam mail, Spammers, Security**

ARTICLE INFO

Article History

Received: 28th May 2016

Received in revised form :

28th May 2016

Accepted: 31st May 2016

Published online :

1st June 2016

I. INTRODUCTION

The spammed emails can be expressed as unbidden and unnecessary Emails sent with the motive of pecuniary profit or hazardous threats to the system. They may be used to distribute viruses or fake announcements that cause responders an average loss of 25 USD per reply. The recently done survey forecast that the 1 user among 50,000 users open or give replies to the spammed mails Moreover, the fact that 58 billion out of the 90 billion emails which are composed or sent daily are the fake mails or the threatened mails this underlines for both the instant and importance of creating and adopting effective processes for received emails. Identification and prevention of the spam is the most crucial work in the pattern recognition and data mining advancement as furious development has been done already for developing algorithms which are capable enough to identify the spam from the fake or the threatened emails. Emails are mainly filtered on the basis of its content in the body, it may include text and photos or videos or other stuff which may provide the detailed information about the sender of the email.

In this paper we have compared some proposed content based filtering algorithms that are based on the text classification for determining whether the mail is threatened or not. The rest of the paper continues as follows. Section II proposes the related work required for this paper mainly content based filtering and policy based personalized of OSN techniques. This paper also provides the details about the existing systems. This paper also presents the proposed system and the future scope in the sections IV and V respectively.

Social networking mediums which are available now a days are the most famous and the easiest way for communicating ,sharing and levigating a vast amount of information about the people's life. Genrally daily and continuous communications implies the sharing of the numerous types of materials which includes texts, photos, audio clips and otther stuffs such as the video materials. According to the recent survey the statistics of facebook and twitter users shows thar the average user's shares approximately 90-95 pieces of data every month The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data. They are instrumental to provide an active support in complex and sophisticated tasks involved in OSN management, such as for instance access control or information filtering. Information filtering has been greatly explored for what concerns textual documents and, more recently, web content. However, the aim of the majority of these proposals is mainly to provide users a classification

mechanism to avoid they are overwhelmed by useless data. In OSNs, information filtering can also be used for a different, more sensitive, purpose. This is due to the fact that in OSNs there is the possibility of posting or commenting other posts on particular public/private areas, called in general walls. Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages. We believe that this is a key OSN service that has not been provided so far. Indeed, today OSNs provide very little support to prevent unwanted messages on user walls. For example, Facebook allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends). However, no content-based preferences are sup-ported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them. Providing this service is not only a matter of using previously defined web content mining techniques for a different application, rather it requires to design ad hoc classification strategies. This is because wall messages are constituted by short text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences. The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. We exploit Machine Learning (ML) text categorization techniques [4] to automatically assign with each short text message a set of categories based on its content. The major efforts in building a robust short text classifier (STC) are concentrated in the extraction and selection of a set of characterizing and discriminant features. The solutions investigated in this paper are an extension of those adopted in a previous work by us from which we inherit the learning model and the elicitation procedure for generating pre-classified data.

II. RELATED WORK

A Content Based Filtering

In content-based filtering, each user is assumed to operate independently. As a result, a content-based filtering system selects information items based on the correlation between the content of the items and the user preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences. While electronic mail was the original domain of early work on information filtering, subsequent papers have addressed diversified domains including newswire articles, Internet "news" articles, and broader network resources. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The activity of filtering can be modeled, in fact, as a case of single label, binary classification, partitioning incoming documents into relevant and no relevant categories. (More complex filtering systems include multilevel text categorization automatically labeling messages into partial thematic categories)

III. EXISTING SYSTEM

Cache architecture consists of two lists namely black list and white list.

Black List: In our research we have put a few domains and email ids in the black list which were presumed of causing danger or threat. For example those websites can be put in blacklist which have a past record of fraudulent or which exploits browser's vulnerabilities. In creating a filter; if the sender of mail has its entry in the black list then that mail is undesirable and will be considered as spam.

White List: It is opposite to the black list concept. It consists of the list of entries which can penetrate through and are authorized. These mails are considered as ham mails and can be accepted by the user. It has a set of URLs and domain names that are legitimate. After creating both the lists when any email arrives the 'To' and 'From' field is extracted from its subject to check if it is in the black list or the white list. The main rule applied here is that if the sender is from the black list then it will be considered as a spam mail.

IV. PROPOSED SYSTEM

A fundamental problem we often study is about leveraging the secrecy of a small piece of knowledge into the ability to perform cryptographic functions (e.g. encryption, authentication) multiple times. In this paper, we study how to make a decryption key more powerful in the sense that it allows decryption of multiple cipher texts, without increasing its size .We Solve the problem by introducing a special type of public-key encryption which we call key-aggregate cryptosystem (KAC). In KAC, users encrypt a message not only under a public-key, but also under an identifier of cipher text called class. That means the cipher texts are further categorized into different classes. The key owner holds a master-secret called master-secret key, which can be used to extract secret keys for different classes. More importantly, the extracted key have can be an aggregate key which is as compact as a secret key for a single class ,but aggregates the power of many such keys, i.e., the decryption power for any subset of cipher text classes.

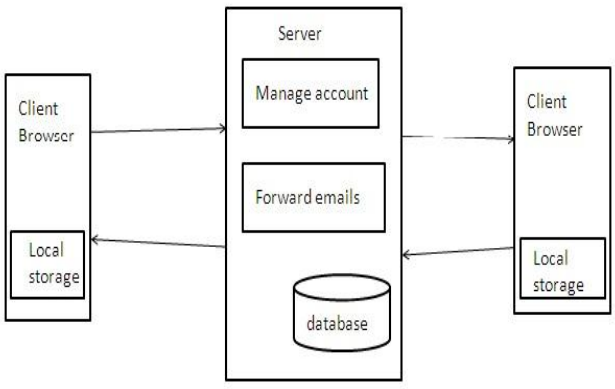

Fig. System Architecture

Module:
Client:
Client menace number of user cans browse the emails. He can perform different activity like manage account forward emails etc.

Server:
Server can perform interface between client browser and database. Server is strong parameter in our project, if server is slow down then the forwarding or receiving emails also slow down.

Database:
Database can stored the all reporting data; user can perform in sending or receiving all types of emails.

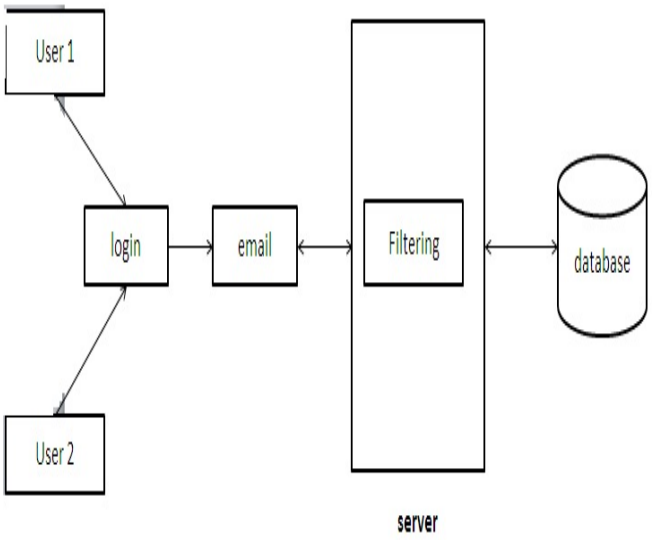

Fig 2. Functionality working

V. RESULT

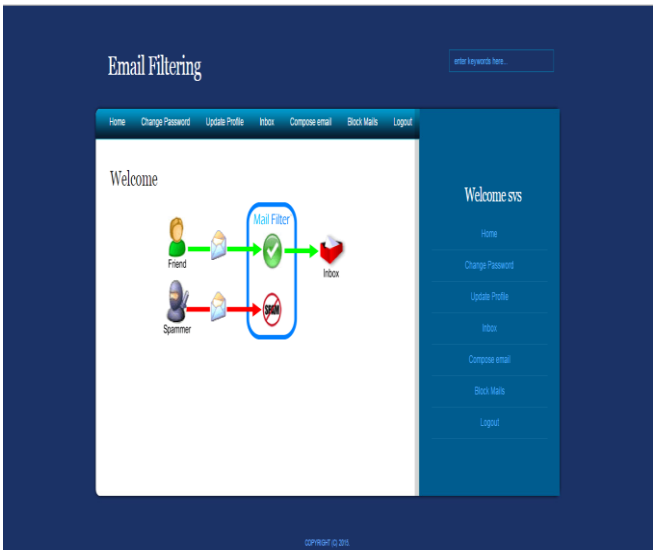

Fig 3. Home page

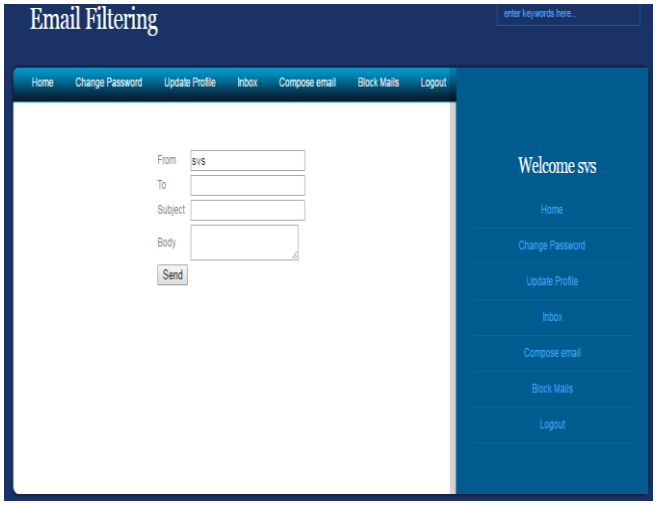

Fig 4. Compose mail

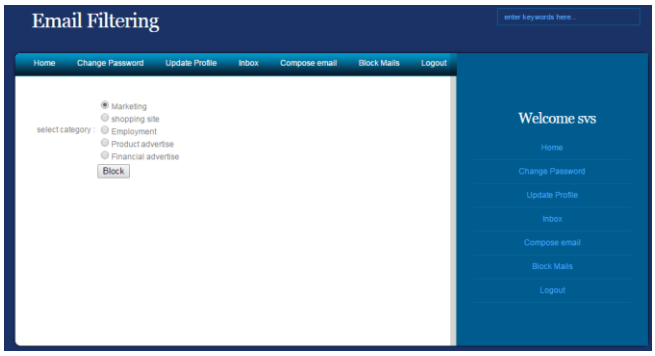

Fig 5. Block incoming mail

VI. CONCLUSION

In this paper we have presented a basic study of the SVM, LM-SVM, Decision Trees, Neural Networks and Artificial Neural Network classifies on the basis of the perfection, accuracy and the performance of the algorithms. The better approach of this paper will be the additional features added for classifying the ham or spam mails using advanced email Filtering algorithms.

REFERENCE

[1] Miszalska, I., Zabierowski, W., & Napieralski, A. (2007, February). "Selected Methods of Spam Filtering in Email." In CAD Systems in Microelectronics, 2007. CADSM'07. 9th International Conference-The Experience of Designing and Applications of (pp. 507-513). IEEE.

[2] Scholar, M. (2010). "Supervised learning approach for spam classification analysis using data mining tools." organization, 2(08), 2760-2766.

[3] Youn, S., & McLeod, D. (2007). "A comparative study for email classification." In Advances and Innovations in Systems, Computing Sciences and Software Engineering (pp. 387-391). Springer Netherlands.

[4] Xiao-li, C., Pei-yu, L., Zhen-fang, Z., & Ye, Q. (2009, August). "A method of spam filtering based on weighted support vector machines." In IT in Medicine & Education,

2009. ITIME'09. IEEE International Symposium on (Vol. 1, pp. 947-950). IEEE.

[5] Sculley, D., & Wachman, G. M. (2007, July). "Relaxed online SVMs for spam filtering." In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 415-422).

[6] Chan, T. Y., Ji, J., & Zhao, Q. "Learning to Detect Spam: Naive-Euclidean Approach." International Journal of Signal Processing, 1.

[7] Puniškis, D., Laurutis, R., & Dirmeikis, R. (2006). "An artificial neural nets for spam e–mail recognition." Elektronika ir Elektrotechnika (Electronics and Electrical Engineering), 5(69), 73-76.

[8] Drucker, H., Wu, D., & Vapnik, V. N. (1999). "Support vector machines for spam categorization." Neural Networks, IEEE Transactions on, 10(5), 1048-1054.

[9] Provost, J. (1999). "Naıve-Bayes vs. Rule-Learning in Classification of Email." University of Texas at Austin.

[10] Medlock, B. (2006, July). "An Adaptve, Semi-Structured Language Model Approach to Spam Filtering on a New Corpus." In CEAS

