A Hybrid and Auto-adjusted Spam Filter

Shu Bin Chen
Department of Computer Science and Communication Engineering
Providence University
Taichung, Taiwan, R.O.C.
g9873007@pu.edu.tw

Hsiao Ping Lee
Department of Applied Information Sciences
Chung Shan Medical University,
Taichung, Taiwan, R.O.C.
ping@csmu.edu.tw

Tzu-Fang Sheu
Department of Computer Science and Communication Engineering
Providence University
Taichung, Taiwan, R.O.C.
fang@pu.edu.tw

Abstract—**The spam E-mail problem has become more and more serious today. Enterprises and users have to spend lots of time on filtering out useful messages from spam. A variety of spam filtering mechanisms had been proposed, including list-based method, behavior-based filter, content-based method, and cocktail filtering mechanisms. In order to improve the accuracy of spam filters, this paper proposes a novel spam detection system, which combines a behavior-based method and a blacklist method. The proposed system will analyze spam behaviors and then generate behavior patterns. Additionally, the propose system will use an auto-updated blacklist mechanism which collects the updates from anti-spam organizations. The proposed system also uses a feedback mechanism that adjusts the behavior patterns based on users' responses. Therefore, the proposed spam detection system can perform more efficiently and accurately.**

Keywords- spam; list-based method; behavior-based filter; content; Internet.

I. INTRODUCTION

Spam (unsolicited bulk mail, or spam mail) was defined as sending a lot of E-mail without recipients' authorization [1]. A large number of spam mails always waste not only users' time but also network/system resource. With the rapidly growing Internet, spam problem has become more and more critical. According to the security report from Symantec MessageLabs in 2010, it says that the average ratio of the global volume of spam E-mail is up to 89.1% [2]. Spam has become a critical research issue. A variety of spam filtering mechanisms had been proposed, including list-based method, behavior-based filter, content-based method, and cocktail filtering mechanisms.

More and more researches apply hybrid filtering technologies to spam filters to obtain higher accuracy rate. However, these methods are more complex than basic filtering technologies and so that result in low computing performance. Moreover, when the weights of the combined technologies are not configured appropriately, the system may perform badly. In this study, considering the efficiency and accuracy of the spam filter, we combine a Behavior-based method and a blacklist method as a hybrid and auto-adjusted spam filter.

This study is divided into four sections. Section I briefs the motive of the study, the faced problems and the goal of the study. Section II shows related background, describing the previous mechanisms of spam filter. Section III explains the system architecture. Section VI describes the expected result of this research.

II. BACKGROUND

A. *List-based filter*

The common list-based methods are based on blacklists and white-lists [6]. Blacklist spam filter blocks the messages based on senders' IP addresses or domains that listed in the database. Some anti-spam organizations release free blacklists, e.g., open relay blacklist. Contrarily, white-list filter only accepts the messages which senders' IP addresses or domains are listed in the database. The list-based spam filters have to update and maintain the lists regularly to keep the lists correct. However, the sender of spam E-mails are mostly fake, and thus the accuracy of list-based spam filter is not good. In general, list-based spam filters usually cooperate with other kind of filtering mechanisms [6].

B. *Content-based filter*

Content-based spam filters define keywords or signatures of the spam, and search the keywords in the E-mail content. While the keyword is found in message content, the message will be marked as spam. Content-based filter are divided into two categories based on the generation of keywords and the decision methods: heuristic filter and statistical filter. The heuristic-based filter defines the keyword manually. The statistical-based filter divides E-mail content into many small tokens, and uses statistical formulas to calculate the probability that the message is a spam based on the tokens. The content-based spam filter has good detection rate. However, content-based filter has to scan the entire message that results in low performance and high latency.

Jiansheng et al. [8] proposed a statistical-based Bayesian filter for spam detection. This method collects a large number of spam and non-spam messages, and then analyzes the messages. After the preprocessing and training phase, two spam token databases are constructed based on the level of damage of spam: ordinary spam database (including commercial and educational) and malignant spam database (including religious, political and pornography). The appearance probability of each token is calculated and stored in three hash tables: hash_good for non-spams, hash_bad for ordinary spam and hash_very_bad for malignant spam. For each incoming message, it has to be divided into small tokens. Then, based on Bayes Theorem and the appearance

 ISBN: 978-1-61208-139-7

probability of pre-sampled tokens in the hash tables, the probability of the incoming message being a spam can be calculated.

Jungsuk et al. [10] used heuristic feature selection method to artificially measure the availability of each feature. This system was composed of several phases. Firstly, it collects E-mail content and then extracts URLs in the content. Then, the URLs are forwarded to a crawler machine and a feature extractor. The crawler machine will download html content based on the URLs' locations. Then a cluster recognizer will calculate the similarity between the extracted URLs and the downloaded html content from the crawler machine, and then recognize spam mail clusters. A feature extractor will extract twelve features from spam mail and the extracted URLs. Finally the selector will use the obtained results to selected significant features. With the selected features, spam can be properly classified.

TABLEI COMPARISON OF COMMON FILTERING TECHNIQUES

Filter technology	Advantage	Disadvantage
List-based filter	Faster execution rate.	High false positive rate.
Content-based filter	Good accuracy.	Must scan full content of E-mails, extremely expensive system resource.
Behavior-based filter	Fast execution rate, fewer network delays, and no regular maintenance database.	High false positive
Anti-spam Cocktail	If properly adjusted, the accuracy rate is the highest.	If use improper weights for the techniques, the system performance will be bad.

C. Behavior-based filter

Usually, spam mails contain similar characteristics. By analyzing the behaviors of spam, some researches find that some characteristics, such as the sending source, sending time, or transmission frequency, can be used to determine whether the sender is a spammer. The behavior-based spam filter can detect spam by just part of E-mail information, instead of downloading full content of an E-mail, which results in low latency. Although behavior-based spam filter does not have to scan full content of messages and thus gets good efficiency, the accuracy of this kind of filters is not well. In order to improve the accuracy, the behavior-based filter often cooperates with other filtering methods.

Meizhen et al. [3] used mining techniques to analyzecollected E-mail log, which includes server IP, frequency, content length, etc. After analysis, the less relevant information will be omitted, and only the high-related features are used to describe the behavior of spam.

Because spam features may be described in form of discrete values or continuous values, the values have to be pre-processed by data conversion and data compression. Data conversion is done by Fuzzy-ID3 algorithm, and then Fuzzy decision tree rule base is obtained. According to the generated rules, the behavior pattern of spam can be understood clearly.

Meizhen et al. [4] observed the behavior pattern of outgoing messages to obtain E-mail features, such as message size, the number of attachment file, etc. Then the authors used these features to create user sending model. They analyzed the E-mails, and found that most spam messages had URLs. Hence URL model was created. The user sending model and the URL model were combined into the Bayesian filtering system to detect spam. The experimental result showed that the detection rate of this research [4] was higher than 85%.

Gert Vliek [9] collected suspicious IP addresses that have suspicious behavior, i.e., a suspicious spammer may send a lot of network traffic, but receive only single message or zero network traffic. To detect a spam machine, firstly the detection system has to find out suspicious machines. Then, the probability that the suspicious one is a spammer is calculated. When the value of probability is higher than a preset threshold, the suspicious machine may be spam machine.

D. Anti-spam Cocktail

A variety of filtering techniques are used together in the anti-spam cocktail filter, i.e., Bayesian filtering technique combines with blacklist to filter out spam E-mails [7]. In general, the cocktail-based approach has lower false positive, but has a big drawback. That is, if the weights of different filtering technologies are not deployed appropriately, the system will perform badly. Additionally, as the system combines different approaches, the cocktail-based spam filter becomes more complicated and expenses more system resources, and leads to relatively slow execution speed.

The spam filters mentioned previously have different advantages and disadvantages. The comparisons are listed in the Table I. Considering the efficiency and accuracy of the spam filter, this study will propose a hybrid spam filter that combines behavior-based and blacklist-based filtering techniques and also user's feedback mechanism achieve efficient performance and accuracy.

III. A HYBRID SPAM FILTER

Using two or more different filtering mechanisms usually achieves higher accuracy than using single filtering technique [5]. Therefore, this study develops a novel hybrid spam filter, combining behavior-based techniques and auto-updated blacklist. Because behavior-based method and blacklist-based method are criticized for their high false positive rate, a feedback mechanism is involved in the proposed system, which keeps adjusting the behavior model to make the system more accurate.

The proposed hybrid spam filter consists of three major mechanisms: a blacklist filtering mechanism, a behavior filtering mechanism, and a feedback mechanism. The blacklist mechanism will automatically renew the black IP addresses or domains by referencing the updates from anti-spam organizations. The behavior mechanism will analyze the collected messages and information to obtain useful features, and based on Bayes Theorem, the probability that

 ISBN: 978-1-61208-139-7

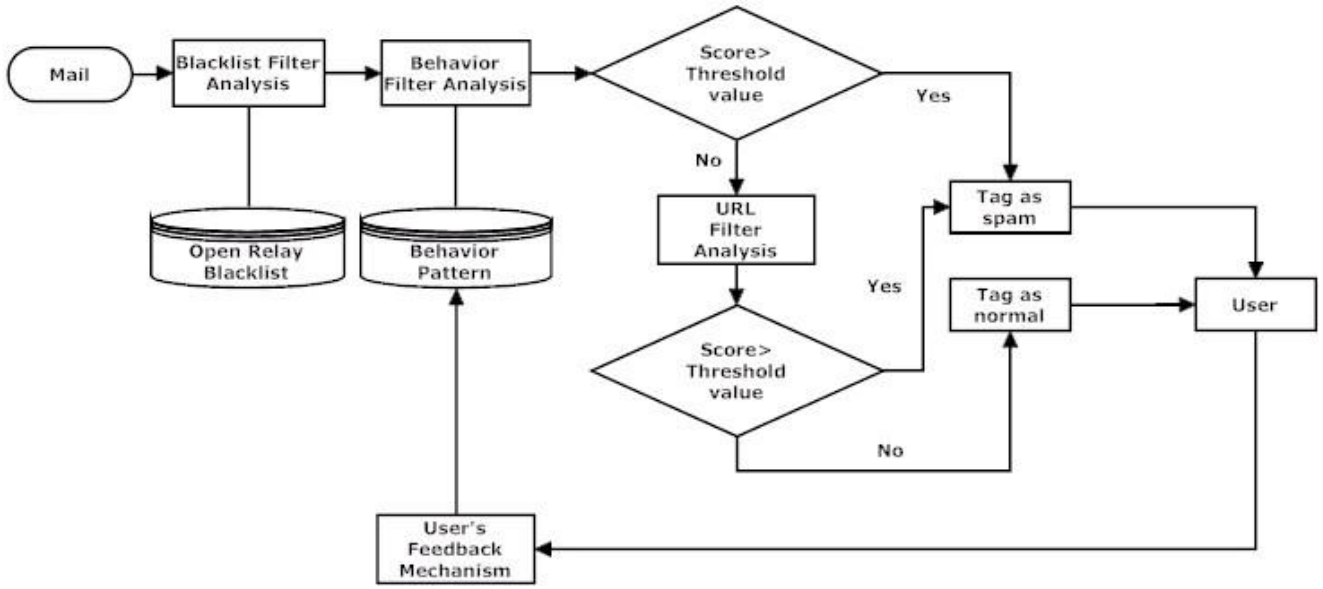

Figure 1. THE ARCHITECTURE OF THE HYBRID SPAM FILTER.

the incoming message is a spam is calculated. The behavior model is automatically adjusted based on users' feedback. Because the behavior-based and blacklist-based models are know as with the advantage of efficient execution speed, and the feedback mechanism can improve the precision, the proposed hybrid filter will achieve better performance.

The hybrid filtering system can be divided into four phases, and the system architecture is shown in Figure1.

Phase 1: With anti-spam organization provided blacklist, filter out the sender which is in the blacklist.

Phase 2: Using behavior model to detect the rest of the mail. Based on the behavior model and suspicious probability, each message will get a score. If the score is higher than a preset threshold, the incoming message is treated as a spam.

Phase 3: For the messages that passes the behavior-based filter, it will be put into a URL filter to check whether the message contains any URL that often considered as a spam URL, to determine whether the message is a spam. If its score is higher than the threshold, it is considered as a spam. On the contrary, it is considered as a normal message.

Phase 4: When users receive the message, they can check whether the decision is correct or not. The system provides a feedback mechanism for the users to respond their corrections. Base on the feedbacks from users, to the system will modify the behavior pattern to improve the accuracy of the filtering system.

IV. EXPECTED RESULTS

This study still works in progress. The proposed system will be implemented based on an open-source spam filter in the Apache SpamAssassin project [11]. The training data will be collected from the Providence University campus network and then used to establish the behavior model. While the blacklist-based filtering technique and behavior-based filtering do not require to scan full message and thus have the advantage of fast detection speed, and the feedback mechanism used in the proposed system can automatically adjust the behavior model to improve the detection accuracy, the proposed hybrid spam filter would have efficient detection speed and better accuracy.

ACKNOWLEDGEMENTS

The authors would like to thank the National Science Council of the Republic of China, Taiwan, R.O.C, for financially supporting this research under Grants 98-2218-E-126-001-MY2.

REFERENCES

[1] The Definition of Spam, http://www.spamhaus.org/definition.html.

[2] Symantec Announces MessageLabs Intelligence 2010 Annual Security Report, http://www.symantec.com/about/news/release/article.jsp?prid=20101207_01.

[3] W. Meizhen, L. Zhitang, and Z. Sheng , "A Method for Spam Behavior Recognition Based on Fuzzy Decision Tree," IEEE, Ninth International Conference on Computer and Information Technology, pp. 236-241, 2009.

[4] W. Meizhen, L. Zhitang, X. Ling, and Z. Yunhe, "Research on Behavior Statistic Based Spam Filter," IEEE, First International Workshop on Education Technology and Computer Science, pp. 687-691, 2009.

[5] Y. Hassan and E. Tazaki, "Rule Extraction Based on Rough Set Theory Combined with Genetic Programming and Its Application to Medical Data Analysis," IEEE, Intelligent Information Systems Conference, the Seventh Australian and New Zealand, pp. 385-390, 2001.

[6] C. Zhi Chen, "A Two Stage Spam Mail Filtering Method Based on Personal Mail," National Taiwan University of Science and Technology, Department of Computer Science and Information Engineering, 2007.

[7] S. Shih Neng, "Design and Implementation of A Personalized Chinese Spam E-mails Filtering System," National Dong Hua University, Department of Computer Science and Information Engineering, 2005.

 ISBN: 978-1-61208-139-7

[8] W. Jiansheng and D. Tao, "Research in Anti-Spam Method Based on Bayesian Filtering," IEEE, Computational Intelligence and Industrial Application, pp. 887-891, 2008.

[9] G. Vliek, "Detecting spam machines a Net-flow data based approach," Faculty of Electrical Engineering Mathematics and Computer Science, 2009.

[10] S. Jungsuk, E. Masashi, K. Hyung Chan, I. Daisuke, and N. Koji, "A Heuristic-based Feature Selection Method for Clustering Spam Emails," 17th International Conference on Neural Information Processing, pp. 290-297, 2010.

[11] Apache SpamAssassin project, http://spamassassin.apache.org/index.html.

 ISBN: 978-1-61208-139-7