Effective Scheme for Reducing Spam in Email System

S. Venkatesh, K. Geetha, P. Manju Priya, N. Metha Rani

Abstract- Nowadays Email is the most widely used communication system that enables the users to transfer their files that may be documents, images, audios, videos etc. Although email system serves for effective communication, it has a large drawback that it is highly exposed to spam. There are various schemes that help to filter the spam mails. But all those schemes provides spam filtering in the receiver side. The disadvantages of using a spam filter in the receiver side are the (1) Wastage of the network Bandwidth required to transmit a spam mail. (2) Wastage of the mail server storage. (3) There is no way to reduce the spam sender’s activities. To overcome the above disadvantages, this paper suggests a method that detects the spam bots and blocks it at the sender side itself, using the bloom filter technique.

Index Terms – Recipient Email Address (REA), Spamming Bots, Bloom filters.

I. INTRODUCTION

A. Overview

Email has been used worldwide due to the large number of efficient services provided by the email system. Email offers fast delivery of mails, compatible environment, efficient file sharing etc. These services made email to be used commercially and in business applications. In spite of the several advantages, email has a large drawback due to the spread of malwares such as viruses, worms etc. through spam mails.

Email service providers provide several security services to ensure authentication and data integrity. These service providers also offer features that filter spam mails. The spam mails are detected and sent to the receiver’s spam folder.

But there are some disadvantages in this existing system mentioned above in the abstract. To overcome those disadvantages, our proposed system gives a system that detects the spamming bots using bloom filters. If some mails are identified as spam then those mails would not be transmitted to receiver i.e., the spam mail is blocked in the sender side itself. The repeated senders of the spam mails are identified by REA technique and are blocked by the system admin.

II. RELATED WORK

[1] (Shuji Suwa; Nariyoshi Yamai; Kiyohiko Okayama; Motonori Nakamura; Keita Kawano; Gada 2012 IEEE)

DSNBL is an effective anti-spam technology but it has a drawback that it cannot filter spam mails in situations when a spammer creates websites using Botnets, fast-flux, Wild Card DNS record. The spam behavior is determined by accessing the DNS record and a black list is prepared for IP addresses that is detected as spam senders.

[2] (Manjusha K.; Rakesh Kumar, 2010 IEEE)

Unsolicited commercial e-mail (spam) is a big threat to commercial users. This paper made an attempt to classify email spam by combining Bayesian network and neural network classification approach. It analyzed the header information like sender details and origin IP etc. using centered Bayesian network, whereas the content and subject of the email were separately analyzed and classified by neural network.


Spammers aims to find the way to avoid filtering out from the spam filters. On the other hand, the email recipient and network system/administrator, aims for effective spam mail filtering technique to detach the spam mails. The problems of spam mail filtering are that each user has different perspective toward spam mails. In this paper, behaviors of spammers are analyzed and classified based on the Random Forest algorithm.

[4] (Sheng Wen; Wei Zhou; Jun Zhang; Yang Xiang; Wanlei Zhou; Weijia Jia; Cliff C. Zou, IEEE 2014)

Modeling the propagation dynamics of email malware is a basic technique for predicting the potential damages and developing countermeasures. Modern email malware exhibits two new features, re-infection and self-start. To Address the problem, a novel difference equation based analytical model is derived by introducing virtual infected user.
III. SYSTEM ANALYSIS

A. Problem Definition

The people create an account on particular cloud e-mail service provider among many of the e-mail service providers. There are some default schemes to provide better services to all mail users. But sometimes mail users want exchange some type of files through e-mail. Because cloud e-mails are accessed from anywhere in the world. If anyone mail user send attachment with any one kind of virus among different type of viruses. After composing mail this mail will reach the mail sever. At mail server if there is no congestion then the mail server checks the mail content, otherwise the mail server will clear the congestion but do not filter the spam mails. In the destination user system, some kinds of virus protection software are used to remove viruses from attachments. But sometime this software wont worked properly, so destination users system would get affected by the viruses. There are chances that viruses affect or crash the mail server.

Limitations:

1. High Protection Software Costs.
2. Time consuming
3. Chances to crash the server
4. Systems or mail network may be affected

B. Proposed System Feature

In this proposed system, the bloom filter technique is implemented at mail user’s side. Now the mail users create a mail then send it. After sending that mail, initially mail body is read by the server, and if the server finds any wrong words in that particular string, those words are stored in the bloom filter dictionary and also mail users suggest some of the words as wrong words. If wrong word is present then server block this mail automatically. On the other hand, the mail server also checks the file attachment. The file attachments are read as the streams and the hash values for each stream in the attachment is calculated. If there is no hash value for some of the streams in attachments then it will be considered as the viruses and the mail will not be transmitted to the receiver. If all of the streams has hash values then there is no viruses in that attachment, and the mail is transmitted to the destination.

Advantages:

1. No Need for Protection Software.
2. Saves network Bandwidth.
3. No chance for server crashing
4. less time consumption.
5. Saves mail server storage.

SYSTEM DESIGN AND IMPLEMENTATION

A. Architectural Design.

The Architectural design gives the diagrammatic view of the working of the proposed system. The proposed system involves two kinds of monitoring:

1. Network monitoring.
2. REA monitoring.
B. Software Design
The software requirements document is the specification of the system.

1. ODBC

Microsoft Open Database Connectivity (ODBC) is used as a standard interface for application developers and database system providers in the programming environment. Before the use of ODBC, windows programs to interface connected with database systems. The programmers used a proprietary languages for each database that has to be connected. The ODBC changed the database systems from a coding perspective to an user-friendly manner.

The ODBC system files cannot be installed into a system with the Windows95. Instead, one can install the ODBC when database applications like SQL server client or Visual Basic 4.0 is set up in the system. After ODBC is installed, the ODBC icon in the control panel uses a file called ODBCINST.DLL. The ODBC data sources can be administered by a stand-alone program, ODBC ADM.EXE. The ODBC has 16-bit and 32-bit version programs that maintains a separate ODBC list of data sources.

2. JDBC

JDBC is developed by Sun Microsystems for Java Database Connectivity or JDBC. JDBC is aimed for setting an independent DataBase Standard Application Programming Interface(API) for java platforms. JDBC offers access to the SQL database for providing a consistent interface to the RDBMSs and those interface is achieved either by using a ‘plug-in’ database connectivity modules or by using drivers. To achieve a JDBC support, one must provide the drive for each platform to run the database and Java.

C. Implementation Setup

COMPOSE MAIL:

In this module a single person can send one or more mails to other person by entering the mail id and subject body and can
also add attachment and then click the send button.

DETECTION OF SPAMMING BOTS:

The spamming bots are identified by the following process:
1. The composed mail subject and the body content is read by the mail server.
2. If any text is identified as an irrelevant word or spam words, then the mail will be reported as spam.
3. The mail attachment is scanned by using bloom filtering technique which takes the stream of contents and calculates hash value for the stream set. If there is any stream that has no hash value, then it will be considered as spam and blocked.

CHAT:

In this Module users can have ‘n’ number of friends from their circle. If any user wants to contact any friend from this circle they can chat / communicate with them through the e-mail communication network.

DETECTION OF ENCRYPTED PACKET:

In this module, the composed mails and chat texts are transmitted to the mail server. For the security reasons, files are encrypted and then split as packets and then transmitted into the network. The server collects the packets and the reassemble the packets to decrypt it.

NETWORK MONITORING:

In this module, the user’s network is monitored by the admin based on the MAC address and IP address. Every user’s MAC and IP are identifiable by the network admin whenever the user accesses the internet. The admin collects the MAC and IP of the spammers based on the spam count.

ANALYZE REA:s:

In this module, the admin collects the mail addresses of the spammers using REA(Recipient Email Address). If there are many number of spam mails transmitted to many users from a particular mail address then that particular mail address is identified and blocked by using REA.
REA MONITORING

Hi Master!

REA Monitoring

<table>
<thead>
<tr>
<th>IP</th>
<th>MESSAGE COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:diwakar@gmail.com">diwakar@gmail.com</a></td>
<td>10</td>
</tr>
<tr>
<td><a href="mailto:mani@gmail.com">mani@gmail.com</a></td>
<td>15</td>
</tr>
<tr>
<td><a href="mailto:neels@gmail.com">neels@gmail.com</a></td>
<td>1</td>
</tr>
<tr>
<td><a href="mailto:rockers@gmail.com">rockers@gmail.com</a></td>
<td>2</td>
</tr>
</tbody>
</table>

SPAM MONITORING:

Hi Master!

Bloom Filter

Click To Network Based

Click To REA Based
VI. CONCLUSION AND FUTURE WORK

A. Conclusion
Whenever a user composes a mail with some specified format and then sending it to another user, the composed mail will be scanned by the mail server before reaching the receiver. If this mail contains any unwanted information or malwares, then it automatically affects destination user system and also affects server system. To alleviate this problem, the bloom filters works on sender side for filtering those unwanted content in the sender side itself. This process leads to secure mail usage among users.

B. Future work
This paper detects the spam mails before it reaches the receiver. The spam detection is based on bloom filters and sentiment analysis. The future work can concentrate on the various anti-spam filtering techniques that detect the spam messages before it reaches the mail server.

REFERENCES

[1] Spam Mail Discrimination System Based on Behavior of DNS Servers Associated with URLs (Shuji Suwa; Nariyoshi Yamai; Kiyohiko Okayama; Motonori Nakamura; Keita Kawano; Gada 2012)

[2] Spam Mail Classification Using Combined Approach of Bayesian and Neural Network (Manjusha K.; Rakesh Kumar, 2010 IEEE)


[4] Modeling and Analysis on the Propagation Dynamics of Modern Email Malware (Sheng Wen; Wei Zhou; Jun Zhang; Yang Xiang; Wanlei Zhou; Weijia Jia; Cliff C. Zou, IEEE 2014)

[5] Efficient Deniably Authenticated Encryption and Its Application to E-mail (Fagen Li, Member, IEEE, Di Zhong, and Tsuyoshi Takagi, IEEE 2016)