**PROJECT PROBLEM INVESTIGATION**

**INTRODUCTION**

Due to the wide popularity of the internet and its communication with no cost, it was recognized as the premium tool for advertising and marketing. With respect to economic constraints, most number of people started sending emails to thousands of people across the world. This made internet, a commercial network with the association of electronic mail as one of the quick resources of communication. The major problem in today’s internet world is sending bulk or unsolicited emails to numerous users. This adds an additional advantage of launching other attacks and wasting of resources [1]. E-mail spam comes under the electronic spam which sends bulk of unnecessary or junk mail of duplicate emails to the recipients.

**ABSTRACT**

Identifying and fixing the affected machines is the key step to resolve any security threats in a network. Because, it becomes a route to launch several attacks such as Denial of service attacks, spamming, stealing user identities and spreading malware etc. Spamming is one of the major threats where attackers perform single attack and make multiple machines in a network as compromised machines. Even though few existing methods like spam signatures and spam behavior analysis resolved the problems to certain extent, it is still not applicable in large networks. Moreover, these methods lack online spam detection mechanism. Existing systems and its drawbacks are also discussed in this report.

**Types of spam:**

* + 1. **Email Spam:**

Email spam is the most familiar spam that most of the users come across every day. Email spam follows three properties i.e., anonymity, mass mailing and unsolicited emails. Anonymity is the property of hiding the uniqueness and whereabouts of the email sender. Mass mailing is defined as the sending of bulk identical emails to the large number of groups and unsolicited emails are the emails transferring to the recipients who do not request. Typically, an email sent to large number of groups without any request by hiding their identity is referred as email spam.

**1.1.2Unsolicited text messages:**

This is kind of similar to instant messenger spam but here the messages are passed

via mobiles. SMS is the service through which the messages are transferred from one

user to other user. The easiest way is to maintain the contact with the known friends instead of strangers. It is relatively easy to find the source where the message is coming from with the instant messenger spam. It is critically important no to click on the links that are passed via mobile by the spammers.

**1.1.3 Social networking spam**: Social networking sites play an important role in today’s world. With the advent of such sites, spammers also started flooding using new techniques to make the social networking sites such as face book, twitter, linked in etc. as part of the spamming activities. As of now it is targeting only the wall posts, messages but these techniques evolve certainly over a period of time. Spammers use notes or messages through various groups or pass the messages with embedded links, which may lead to pornographic or other sites and target spam [4]. Even though

**#Blank spam emails and forwarding spam emails:** Spammers also use the technique of sending a blank email to the recipients. The purpose of sending this type of email is to recognize whether the recipient possesses a valid email ID or not. If it is an invalid email ID, then it bounces back stating with a non deliverable notice. This helps the spammer to identify whether it is a valid email ID or

not. Sometimes blank emails also attach few files that initiate Trojan virus if it is opened

in the system. Forwarded emails are again one more cause to initiate spam emails. This

makes users forcibly to forward the users in their friends list. As a result it forms a chain

and delivers spam emails and becomes uncontrollable. This eats away lots of time and

space and costs a lot to filter spam in the system. [19]

**1.3. Email spam filters:**

Spam filter is a piece of software that is used to filter the spam emails based on the

content and rules adhered by its corresponding software. Every single spam filter has its

own set of rules through which the spam is filtered from spreading across the network. It

involves the content of the spam, address of the users and where it is redirecting to etc.

Based on these parameters it judges, whether an email is a spam or not. There are

multiple spam filters divided based on their rules [21].

**Disadvantages:**

Even though this frame work is accessible for the URL groups embedded in

email, it lacks in action taking for IT security groups. The degree of burstiness in the real

time is one more major drawback that is faced by this frame work. It does not provide

any clarity about how well it is performed in a real time for the spam campaigns.

**2.1 Problem statement:**

The reasoning for choosing the spam filter is that so that lots of malicious mails can be filtered and used for the protecting us from dangerous flaws in the system which can corrupt the system .Nowadays, Spam filter is one of the most important app which can save the lots of organization information from getting corrupt. As the Hackers are more advanced nowadays,so we need to make the Spam Filter more powerful so that we can find the real vs malicious content. I would like to train it with Machine Learning Model so that we would be able to reach to the full extent.

**3.1Characterizing botnets from Email Spam Records:**

This framework presents techniques, which use traces of spam email to map

botnet in groups. This is done by viewing for several bots involving in the same spam

campaign. This has been used against a sample of spam email from Hotmail web mail

system and has successfully detected multiple botnets. This technique uses a large set of

spam emails as input, which are destined at Hotmail in a regular period. Group of botnets

that involve in launching spam emails with respect to its statistics like dimensions and

events are observed in the output. Three major steps are involved in identifying them.

With “clustering email messages,” spam email messages with identical content are

transferred from the entity which is controllable [12]. Firstly, the spam campaigns are to

be identified. Most of the spam content in the campaigns appears to be similar, but in

order to evade the differences; identical properties are to be considered. For efficiency,

shingling concept is used [2]. A unique property like finger print is considered here and

counts the number for every single message. Based on the number of finger prints that

share among the spam campaigns identical nature is computed.

**3.2 Objectives :**The objective of this research project is to design and implement a tool to detect the spam attacks in a network.

Other objectives:

• To critically review spam characteristics in a network and analyze the current

trends in protecting the spam inside a network.

• To identify the loop holes in the existing systems and propose an efficient tool.

• To design and develop proposed tool that identifies spam communications and

keeps track of IP addresses of the machines in a network.

**LANGUAGES AND TECHNOLOGY USED**: Python and SQL server(for connection) and Machine Learning Models(Linear Regression Model, SVM Model and Polynomial Model).

**4.1 Functionalities**:

The flow of execution is explained in terms of different modules. This project is

implemented in .net frame work. The back end is developed in Oracle SQL. All the

modules and the transferring and receiving of data is explained clearly in this section.

**4.2.1 Master Server:**

Master server is the component that is used by the network administrator to keep

track of clients and spam details. This one has its private credentials for the network

administrator to login into this application. It has number of functionalities like adding

clients, sending data, receiving data and generating spam reports. [18]

**4.2.1.1 Add clients:**

Master server can add any number of servers and clients. Once the server is

selected, any number of clients can be added to their respective server based on its IP

address. It has the options of adding client address based on the IP address. An IP address

is also validated when it is specified. Every single client is provided with its login

credentials as well. User ID and password are assigned for all the clients to maintain

personal mail boxes to send an email.

**4.2.1.2 Add server:**

Master server has the functionality of adding any number of servers. Any number

of clients can be added under these respective servers. Servers are added based on the

systems IP address. It keeps track of the number of clients fall under each server.

**4.2.1.3 Spam report**:

This is a very important component that is set by the master server where all the

spam reports are maintained. Based on the content of the emails that are exchanged

between the clients, the spam is recorded and there by a spam report is generated. It

filters out the compromised machines and non-compromised machines based on the spam

it is sent.

**4.3.1 Client module:**

The functionalities of client modules are clearly explained in this section. As

discussed in the above section, the server is capable enough to handle any number of

clients.

**4.3.2.1 Send data:**

For every client, there exist mail boxes through which data can be sent. The

exchange of messages takes place in the form of an email. It has the options of

composing an email through which information is passed on from one client to other

information such as client’s name, IP address, date and time stamps are stored in the

database.

**4.3.2.2 Receive data:**

When a client transfers an email in form of a text, it reaches to the inbox of

another client machine. The mail boxes of clients are similar to that of the general email

set ups and have an option to compose, inbox and sent mail. The client has also the

feature to forward and delete the emails.\

**4.3.3 Spam report module:**

This is the most important module in this project. This module runs under the

server software. It maintains the record of all the spam and non-spam messages, thereby

identifying compromised and non-compromised machines in a network.

**TESTING :**

Testing is a very important module in the software development to verify, validate

and provide quality and service for different components of software. It is used to

minimize the risks by efficient use of resources in the development life cycle.

Following are the test cases of this application validating the client’s credentials

and IP addresses. This tool was also tested using various emails containing spam and

non-spam content.

**CONCLUSION & FUTURE WORK :**

Due to enormous usage of internet technology, there is a huge increase in the

network attacks. Among them, spam is considered as one of the main attacks in launching

various attacks like stealing user identities and spreading malware etc. In this project, a

spam detector is developed, which can monitor and detect the machines involved in spam

across the network. This tool is based on a spam filtering algorithm that has the efficiency

of detecting high percentage of spam. It can differentiate spam and non-spam affect

machines in a network of any size. To avoid network administrator to view non-spam

emails and to maintain the privacy among the clients in a network, encryption technique